mapping science

overlapping science and design

juan f. burgos-chaves

Thesis document submitted in partial fulfillment of the requirements for the degree of:
Master of Arts in Design.

Approved by the MFA Review Board of the Dynamic Media Institute
at the Massachusetts College of Art and Design in Boston.

May 2008.

Brian Lucid
Thesis Advisor
Associate Professor of Design
The Dynamic Media Institute
Massachusetts College of Art and Design

Jan Kubasiewicz
Second Thesis Advisor
Professor of Design
Director, *The Dinamic Media Institute*
Massachusetts College of Art and Design

Joe Quackenbush
Assistant Professor of Design
The Dynamic Media Institute
Massachusetts College of Art and Design

dedication

My thesis is dedicated to my brother Carlos, for he has taught me the biggest of the lessons, the simplest to accept, and the hardest to undertake. He has showed me that all the things I thought important in life really are. He has given me the example and support that I needed to finish this task.

Gracias mi hermano querido.

acknowledgments

I would like to thank ***Brian Lucid*** for his advice and encouragement during my journey at the D.M.I. Thank you Brian for being an inspiration every day and reminding me that passion is the right instinct to follow. Your passion for teaching is like an open book that only has good quotes.

I would also like to thank ***Jan Kubasiewicz*** for his enormous contribution to my personal and professional life, I consider you Jan an example to follow and one of the best teachers I have ever had.

Many thanks to ***Joe Quackenbush*** for his patience and perseverance. Those can only translate into improvements. Thank you Joe.

Special thanks to all my friends at the Office of Continuing Education: ***Anne Marie Stein***, ***Stephen Clark***, ***Jonathan Rand***, ***David Carder***, ***Nancy McCarthy***, ***Susan Mendez-Diez***, ***Joe Doucette***, ***Liz Rudnick***, and ***Sheryl Pace***; the things I would like to say about all of you are too long to be put in this document. All of you gave me the opportunity to feel like home, I have learned tons from each of you. I consider all of you my teachers, mentors, and close friends. Thanks to ***George Creamer***, specially for the laughs but also for all your help since my very beginning at MassArt. To ***Nadia Savage*** who was always there answering weird questions.

Thanks to my father for keeping an eye on me, for being my best friend when I needed it and my greatest model, and for never leaving me. You are in my memory every single second, I love you with all my heart.

To my mother who knows just what I need to be happy and reminds me all the time that it is never too late to say I love you. Mom, having you is what keeps me awake and wanting to be better. Gracias mamita de mi corazón.

To my sister Ariadna for being an example to follow, a great friend, and a light during those dark moments in Boston.

To Mariana, who has learned how to love me and understand me. She makes me give my best and supports me no matter what. Mi Mani, you are the best that could have happened to me.

To Ignacio for the bike rides, the nights and days of productive chatting, and the greatness of having a new brother that I never asked for. The way you manage to be the best professionally and personally inspire those that surround you.

To Carlos for giving me the opportunity to be called your friend, it will always be great having people like you as friends, which by the way are not easy to find.

Jesse and Isola, you opened your home, let us be part of your lives, and fit ours perfectly. We will miss you so much.

Special thanks to Jonathan Reichard and Jesse Spence for kindly letting me use their voices in my prototypes. Jonathan is the voice of the Sequence Generator, and Jesse the voice of Shapes.


CITY

abstract

Mapping Science documents the different steps in the elaboration of prototypes of interactive visualization tools for scientific data manipulation. The thesis focuses in the interaction between scientists and their data, and in the application of design principles to achieve better results when scientists organize, structure, input, store, and visualize scientific data.

Mapping Science relies on a Ph. D. thesis on bat behavior to assess its needs and to study scientific thinking processes that are later compared to those used by visual thinkers. The goal is to demonstrate the importance of visual thinking processes in the creation of hypotheses and the exploration of data.

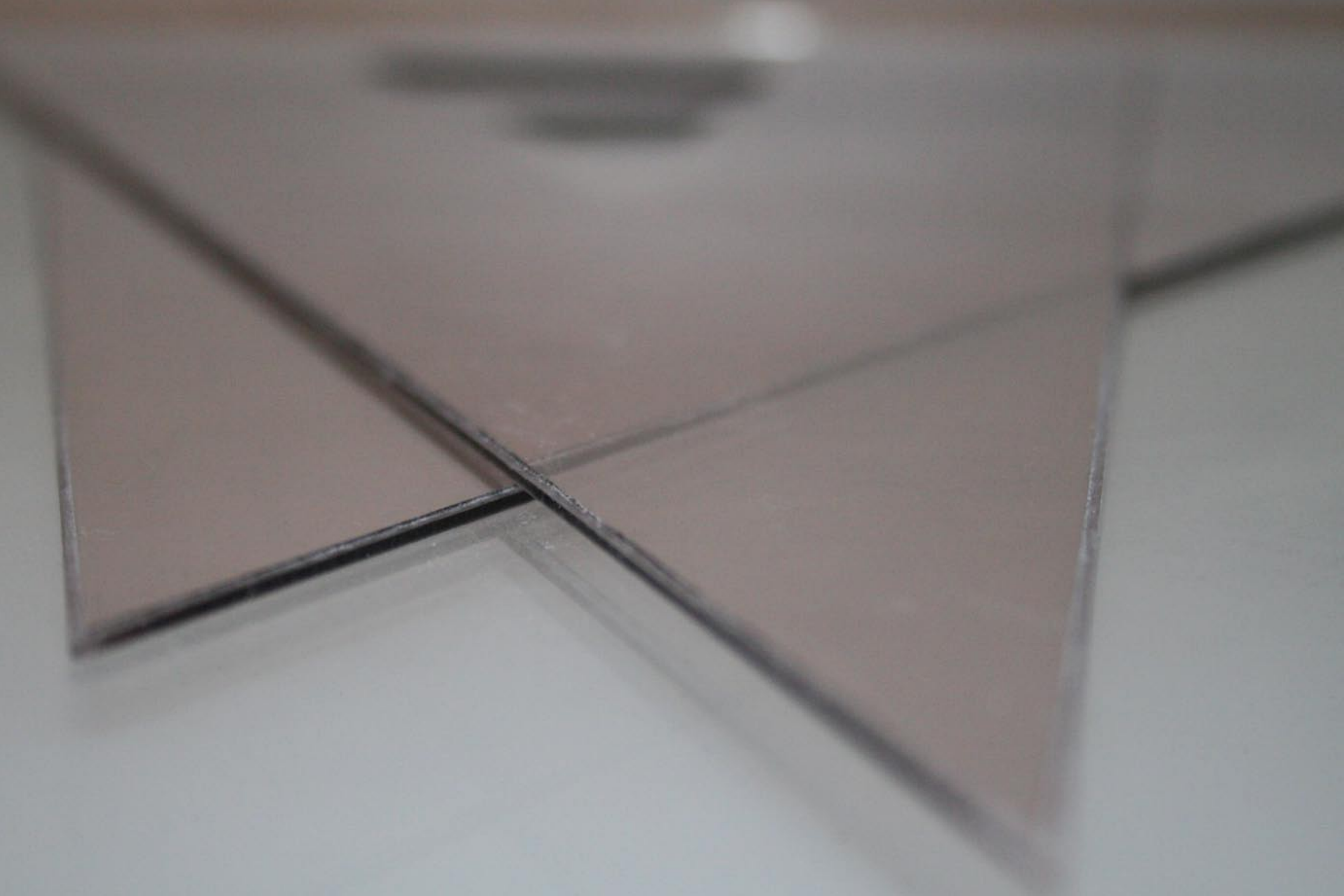


TABLE OF CONTENTS

"The most profound questions of our existence cannot be answered through a mere collection of concrete evidence; at some point, whether we are theologians or automobile mechanics, dentists or draftsmen, each of us reaches a border of the verifiable world, and every one of us leaps. A great deal of what we know, we know only through our imagination—and that knowledge is crucial to our lives."

Peter Turchi

1 / INTRODUCTION

introduction

Disciplines in almost every field of knowledge have always relied on visualization as a tool to communicate their content. Ideas are better explained and easily understood when they are presented visually. For centuries, visualizations of all kinds have set a common ground for audiences and experts to share the primary theories and general information packed in their professional environment.

Major examples of these visualizations are maps and illustrations. Maps are visual representations of a terrain or a three-dimensional space that allow their users to have a better sense of what that space looks like, what are its characteristics, and how users can interact or navigate within the represented surroundings.

Illustrations help us to evaluate in detail properties like form, color, structure, and patterns of the represented object. In some cases we trust graphic illustrations rather than verbal explanations because they are tangible, and let decisions to our own interpretations and not to those of others. Maps and illustrations are just two examples of a long list of visual representations that humans use in their search for efficient communication, but they depict very well the power of images as accessories of this communication. Diagrams, charts, symbols, photographs, and timelines are other examples that fit well the definition of visual representation.

Trail map of the F. Gilbert Hills State Forest. Current times.

To see the picture even clearer, imagine yourself describing with words a map that represents complex patterns, colored routes, and intricate wavy roads. Put yourself in the position of giving directions to a person that doesn't have the visual aid. You could spend significant time trying to describe the map with no success. As detailed as a verbal explanation might be, there will always be space for misunderstandings due to misinterpretations of what you said. It is probable that after a while you start drawing lines creating your own map of the region, allowing the other person to come to a better understanding.

More than 80% of the information processed by our brain is perceived through our eyes, so the need for visual representations attends to the nature of our perception. Not only we do perceive more information with our eyes than with any other sense, but also this information is perceived faster and retained longer than information perceived through other senses like touch, smell and hearing.

William Harvey. Illustration for studies in blood circulation. 1628
The scientific Image. Robin, H. (1992).

The best example of the symbiosis between visualization and fields of knowledge might be that of sciences and visualization. In the case of formal sciences, geometrical representation is a great model. How to understand the size, shape, and position of geometric figures in space without visual representations? Geometry's graphic nature served as the foundation for man's comprehension of the potential in visual explanations. The strongest connection exists between natural sciences and visualization, due to the use of this visual method in scientific problem solving.

Medical sciences have accomplished amazing results in the use of 3-D imaging to create faithful visual reproductions of the human body. Similarly, detailed animations and digital reconstructions allow scientists to understand the physiology and morphology of mammals extinct a million years ago.

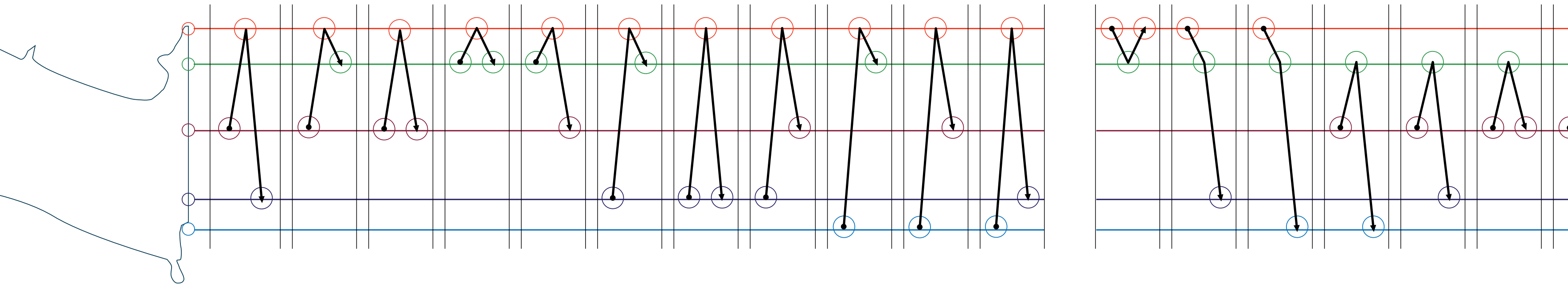
On the other hand, one might see that interesting visual approaches could be ruined due to a lack of knowledge regarding the medium, compositional principles of design, presentation methods, and data organization. Diagrams and graphics created by scientists very often carry all the complexity of their analytical thinking, but do not translate efficiently to images, making those images difficult to read and decipher. Therefore, it is important to understand the need for close collaboration between scientists and designers, complementing the strength of both fields.

Edward Tufte writes in his book "The Visual Display of Quantitative Information":

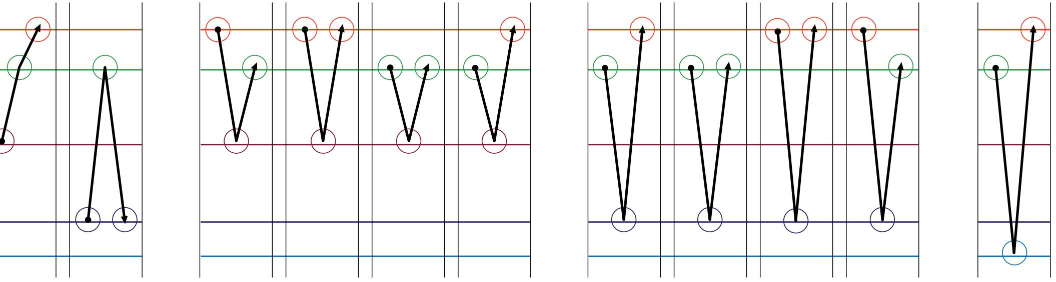
> "Lurking behind the inept graphic is a lack of judgment about quantitative evidence. Nearly all those who produce graphics for mass publication… have had little experience with the analysis of data".

If this is true, we should also postulate that scientists often disregard the need for design solutions to their visualizations. This is why *Mapping Science* will allow scientists to benefit from the experience of visual thinking processes and designers to better understand the process of analytical thinking.

Mapping Science compacts information derived from scientific research, studies its core form, and structures information into visual elements that facilitate its understanding, presentation, and manipulation. The basic outcomes are static and dynamic diagrams, more like maps that seek to represent the scientific territory in which the specific research operates. The same way in which traditional maps give us scaled information of territories, *Mapping Science* diagrams function as graphical representations that display data's hidden peculiarities. *Mapping Science* is a microscope that reveals properties present in data but not visible at a normal scale or state.



2 / TIMELINE / some decisive events



timeline

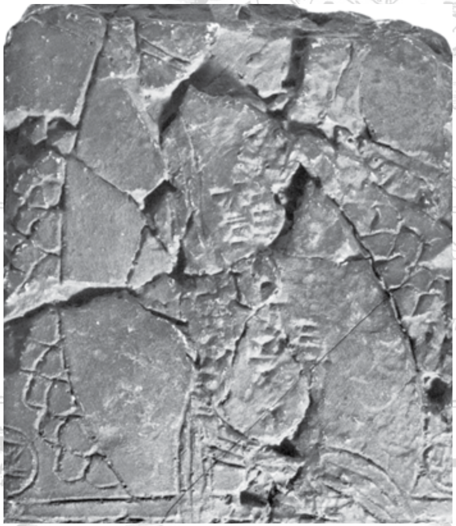
In order to understand the different aspects of the evolution of visualization and its consequences in communication, I offer a sequential placement of events and important precursors of visualization to aid the understanding of the development of this discipline and its presence in time.

The elements displayed in this timeline are only fractions of the complex fields they belong to, functioning as simple start points that were subsequently transformed into complex and vast areas of study. My intention is not to provide a detailed research on these elements but to highlight their important relationship with visualization in general and more specifically, with scientific visualization.

maps

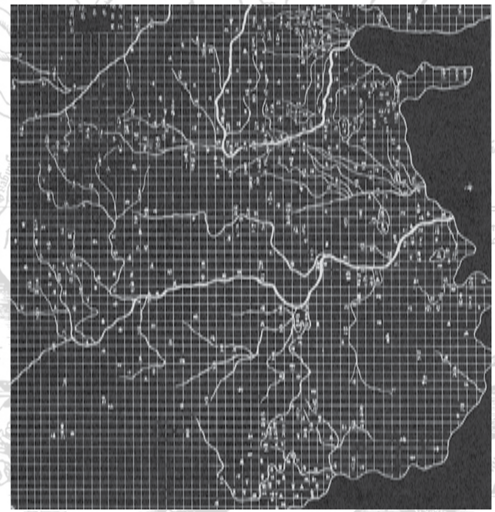
Maps are the most ancient examples of graphical depiction of space and organized information. Precision in scaling our world's surface and efforts in synthesizing the information so it could be interpreted effectively appeared with the birth of maps. There are innumerable examples of other primitive forms of visualization attempting the transference of knowledge or experiences but none achieved the remarkable success in organizing and depicting reality as maps.

At the present time, maps are no longer the flat orientation instruments used to navigate to destinations. The word map implies a much broader concept not only linked to geography but to much more broad applications, such as the representation of mental images as in mind maps, signs, and diagrams. Undoubtedly, the current concept of map is closer to the concept of visualization, since it refers to the act of mapping as in measuring and visually depicting characteristics of anything that can be considered a surface, not only physical surfaces, but conceptual surfaces as well. We can map interactions, data, memory, textures, and almost everything measurable.



Ancient Babilonian Map.
The earlies known map, this clay tablet was part of a cadastral survey, a river runs from the bottom to the top flanked by mountains. Found at Nuzi near Kirkuk in northern Mesopotamia. From the dynasty of Sargon of Akkad.

From: *Maps. Michael and Susan Southworth.*



The *Yu ji tu* or the *Map of the Tracks of Yu*. Extraordinary representation of river systems. Its accuracy is amazing for the time.

From: *Visual Explanations. Edward Tufte.*

ca. 2500 B.C.

3600 years

ca. 1100 A.D.

An example of contemporary improvements to maps are the multi dimensional uses of GPS systems, which are implemented in several applications such as route tracking, map drawing, navigation, measurement of height levels, etc.

The Hereford Map, combines actual geography with legendary, historical, and religious elements. By richard of Haldingham. Decorates the altar of Hereford Cathedral.

From: *Maps. Michael and Susan Southworth.*

ca. 1275 A.D.

illustration

Illustration appeared as the first manifestation of human's necessity for visual images. Dating ca. 20.000 B.C., cave paintings represented complex relationships between the characters depicted and certain features of the images, such as proportion and form, were carefully represented to convey messages and transfer knowledge about hunting techniques or other simple activities.

Illustrations developed by the Egyptians c. 1290 B.C., are evidence of the representation of information at a higher level. By combining elements such as astronomy and mythology, Egyptians illustrated intricate pieces of information leading to the first attempts of mankind to create a written language.

Remigia Cave in Spain. This cave painting shows one of the first attemps of men to visualize interactions with the environment.

From: *The Scientific Image. Harry Robin.*

Ceiling painting in the tomb of Pharaoh Seti I. Astrological, historical, and communicational information together.

From: *The Scientific Image. Harry Robin.*

ca. 20.000 B.C.

18710 years

ca. 1290 B.C.

The important contributions of illustration to the field of scientific visualization began after c. 1.200 A.D., when engravings and woodcut prints were done to illustrate the rapidly increasing body of knowledge. From explanations on how machinery worked to drawings presenting proofs that the earth was indeed a sphere, illustrations became a new way of expression and explanation for difficult to convey information.

Illustration became an important help for the earliest expressions of science: medicine, geometry, botany, and astronomy were some of the scientific fields that used illustration and still use it at present time. With the birth of written language, illustration gained even more importance and the use of both mediums as complements of visualization marked the beginning of visual languages being considered as more than a secondary language.

William Gilbert's Woodcut. As a physician, Gilbert was curious abaut scientific matters beyond medicine. This woodcut shows his interest in magnetism.

From: *The Scientific Image*. *Harry Robin*.

ca. 1600 A.D.

timelines

Many illustrations can be considered timelines since they arrange elements in a chronological order. Egyptians carved their hieroglyphs in walls with a strong sense of chronology and drawings that needed to represent processes displayed step by step explanations, which can also be thought of as chronological visualizations.

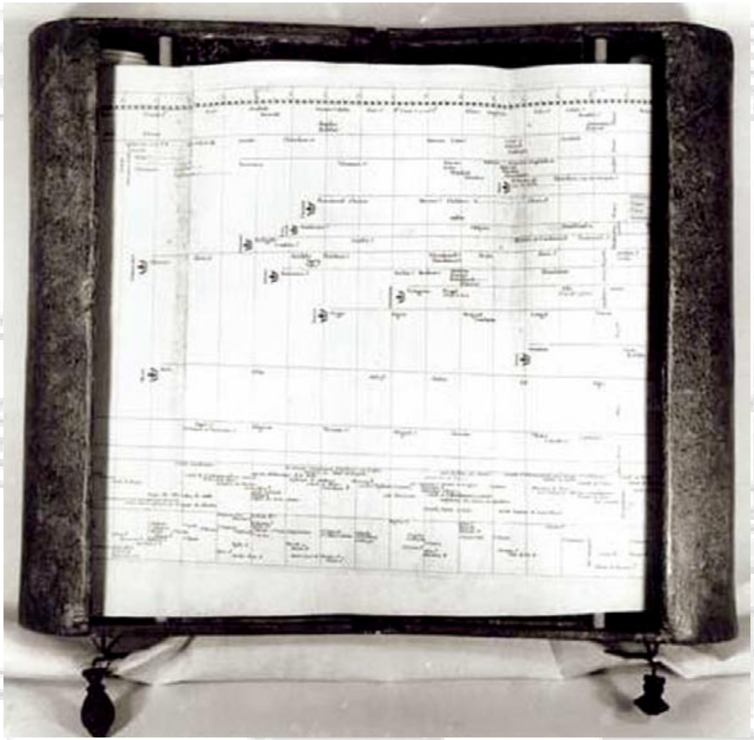
The real concern about how to represent time in diagrams and illustrations appeared during the 18th century, when the passion for the development of knowledge was at its height. The study of complex events and the influence of those events in the society of the time became a principal interest for scholars in the 18th century.

Considered the first modern timeline, the "Carte chronologique" created by Jacques Barbeu-Dubourg (1709-1779) in 1753, was a 54-foot scroll annotated history timeline. In this innovative solution, the format adapted to the need for horizontal placement and thematic categorization of the events.

In science, representation of processes has been seriously influenced by timelines. For example, statistical graphs carry the footprints of timelines, as very often the horizontal axis is used to represent time.

The Carte chronologique: An annotated time line of history. By Jacques Barbeu-Dubourg (1709-1779). Considered the first modern timeline.

From: *Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization. At: http://www.math.yorku.ca/SCS/Gallery/milestone/sec4.html*



ca. 1600 A.D.

graphs

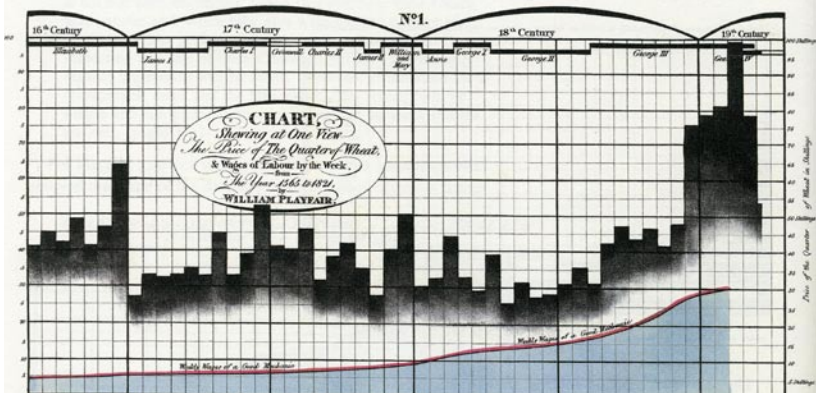
The term graph is another concept with broad applications. Its use and definition varies among different fields, however, graphs remain strongly associated to mathematics, computer science, and, of course, statistics.

Graphs are perhaps, the first visual solution that was created specifically to be applied to scientific data. After the appearance of statistics, there was the need for representation of results, which eventually gave origin to graphs.

Graphs were intended to facilitate the interpretation of data and its relationships. The strong connection between early statistical graphs and the way we currently represent data is obvious. In fact, statistical graphs have only began to evolve recently, as the arrival of computers and the use of dynamic media have given new possibilities to the production of data visualization. For example, the appearance of three dimensions in diagrams and timelines was seen first in the use of the X, Y, and Z axes present in graphs. Computers have facilitated the development of 3-D appearance in the 2-D surface of the screen.

One of William Playfair's statistical charts. Excellent examples of intelligent display of data.

From: *The Visual Display of Quantitative Information. Edward Tufte.*



ca. 1780 A.D.

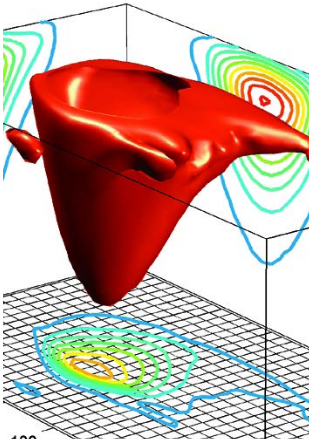
computational graphics

Computational graphics refers to the implementation of computer algorithms to produce visualizations and graphics, or the digital manipulation of visual images to represent data by the means of a computer.

Computational methods for the production of visualizations have evolved in the last sixty or seventy years. Regardless of this relative short time, technological advances, industrial and mechanical innovation, and scientific discoveries have made of computers and computational products a galloping industry.

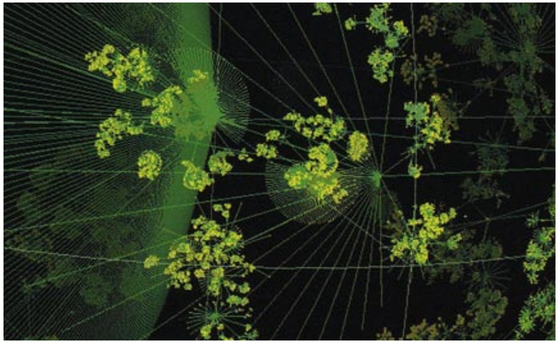
Information visualization has changed more in the last twenty years than ever and computers perfectly fulfill the needs of data and information. Computer algorithmic instructions fit the characteristics of data and serve as a suitable medium to store, display, and organize information.

The potential application of computer graphics is baffling, ranging from natural sciences to social sciences and it is fair to say that there is no current field of knowledge that does not use computer graphics to enhance their content or productivity. Computer imagery in medicine, chemistry, physics, and biology, 3D rendering in all sorts of areas, 2D diagrams created in computers for books and periodicals, computer animations for presentations or for educational purposes, are evidence of the crucial role that computer graphics play these days.



Visualization created using the computational application MATLAB.

From: *Blog. At: http://pouyakondori.blogspot.com/2008/01/introducing-matlab-programm.html*



Visualization of large graphs displaying information of links visited in web pages. Large amounts of data are visualized through this system called Walrus.

From: *Digital Information Graphics. Matt Woolman.*

ca. 1978 A.D.

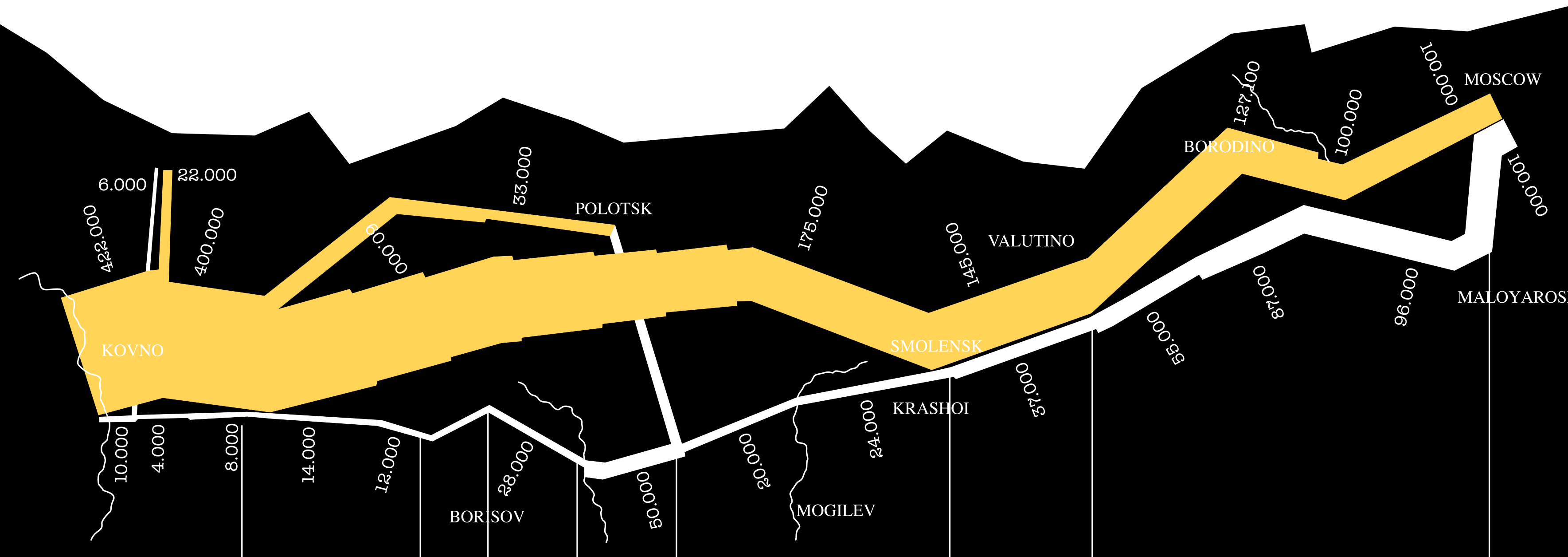
24 years

ca. 2002 A.D.

"... clarity and excellence in thinking is very much like clarity and excellence in the display of data. When principles of design replicate principles of thought, the act of arranging information becomes an act of insight".

Edward Tufte

3 / PIONEERS OF VISUALIZATION



Claudius Ptolemaeus (Ptolemy)

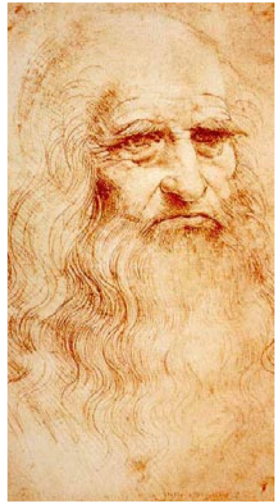
Claudius Ptolemaeus.
b. ca. 83 A.D.
d. 161 A.D.

Image from: *Wikipedia the free encyclopedia.*

With his treatises, specifically the Geographia treatise, Ptolemy pioneered the use of visual representations to improve geographic knowledge of his time. His maps put huge amounts of information aesthetically and wel structured.

Image from: *Wikipedia the free encyclopedia.*

Leonardo da Vinci



Leonardo di Ser Piero.
b. April 15, 1452
d. May 2, 1519

Image from: *Wikipedia the free encyclopedia.*

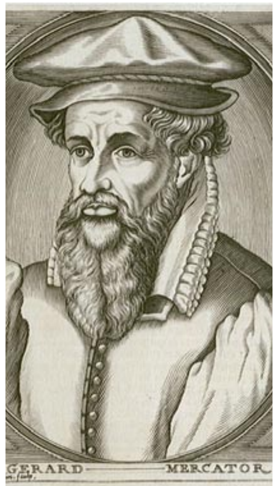
Leonardo's contributions and influence to science were countless and extremely important, by combining arts and science, he changed the future of scientific visualization.

Image from: *Wikipedia the free encyclopedia.*

Gerardus Mercator

Gerard de Cremere.
b. March 5, 1512
d. December 2, 1594

Image from: *Wikipedia the free encyclopedia.*



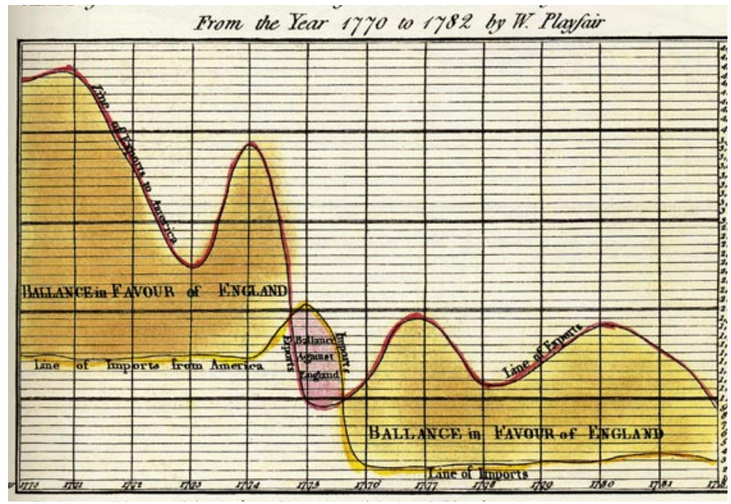
Mercator's maps went way beyond the two-dimensional surface of paper. His contributions to systems of geographical projections are still used today.

Image from: *Wikipedia the free encyclopedia.*

William Playfair

William Playfair.
b. September 22, 1759
d. February 11, 1823

Image from: Image *not available*.



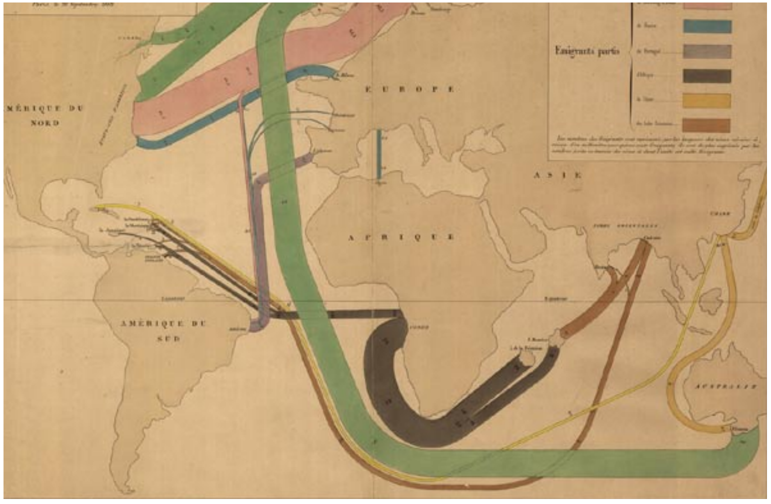
Pioneering presentation of statistical data Playfair created a precedent in visual representation of data. His graphics demonstrated how important visual aid is to explain quantitative data.

Image from: *The Visual Display of Quantitative Information. Edward Tufte.*

Charles Joseph Minard

Charles Joseph Minard.
b. March 27, 1781
d. October 24, 1870

Image from: *Image not available*.



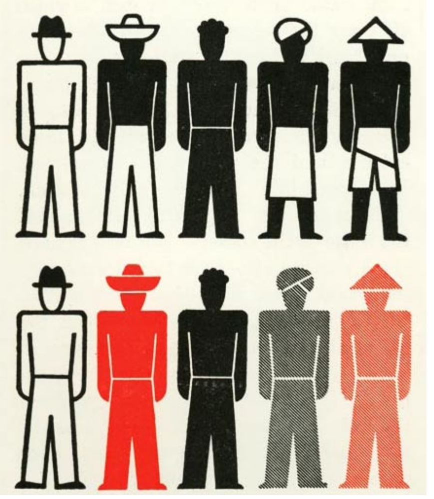
Minard is famous for his beautiful and perfect visual representations of data and information. Many theories and studies have been derived from his graphics.

Image from: *http://www.mapsandglobes.net/19thcentury/maps*

Otto Neurath

Otto Neurath.
b. December 10, 1882
d. December 22, 1945

Image from: *After Neurath Project. At: http://www.stroom.nl/activiteiten/manifestatie.php?m_id=4653044*



Neurath's design of the International System of Typographic Picture Education (ISOTYPE) marked an important date in communication of information without the use of words. Synthesized icons were used to express meaning for statistical and graphical diagrams.

Image from: *Cabinet online. At: http://www.cabinetmagazine.org/issues/24/pendle.php*

4 / SCIENCE AND DESIGN

science and design

A primary concern of my thesis is the complementary role of designers and scientists in the production of visualizations that communicate scientific content, and the importance of interactions between scientists and their data for the effectiveness of these visualizations in the creation of hypotheses and improvement in the research process. Studying how thinking processes for these two fields work allow the participants to have an approach to mechanisms that could facilitate interaction with the parts that conform it.

Connections between scientific and creative thinking have been studied and analyzed for centuries, and many conclusions have been obtained by these studies. It is clear that there are as many similarities as there are incongruities between both processes of thinking. Scientists seem to have problems when using some of the creative methods to find solutions and designers do not find easy to produce their work by utilizing analytical thinking.

What makes the interaction between scientists and designers so difficult is the fact that our brains spend most of the time processing information in either way. The capacity of processing information by the implementation of different methods is directly proportional to the training that we receive.

As in all fields and areas of knowledge there are exceptions to the rule, as there are multiple designers that master visual thinking processes as well as analytical thinking and vise versa, scientists that can produce visualizations with a perfect sense of aesthetics and data structure. However, majority of the cases do not fit into these exceptions and visualizations produced by scientists lack of organizational elements that could have been provided by a specialist in visual communication. In the same order, visualizations produced by artists or designers usually need the expertise and knowledge of the data analysis that only those scientists working with the data have.

A definition of science

"Science is not merely a collection of facts, concepts, and useful ideas about nature, or even the systematic investigation of nature, although both are common definitions of science. Science is a method of investigating nature—a way of knowing about nature—that discovers reliable knowledge about it. In other words, science is a method of discovering reliable knowledge about nature… Science is the only method that results in the acquisition of reliable knowledge… Reliable knowledge is knowledge that has a high probability of being true because its veracity has been justified by a reliable method". Steven Schafersman (1997).

A definition of design

"Design is a set of fields for problem-solving that uses user-centric approaches to understand user needs (as well as business, economic, environmental, social, and other requirements) to create successful solutions that solve real problems. Design is often used as a process to create real change within a system or market. Too often, Design is defined only as visual problem solving or communication because of the predominance of graphic designers. In other fields and contexts, Design might only refer to Fashion Design or Interior Design. However, a recognition of the similarities between all design disciplines shows that the larger definition for Design operates at a higher level and across many media". Nathan Shedroff. (Electronic source)

Early discoveries, even in the area of science, dealt with proportional amounts of visual observations and scientific experimentation. Mark Polland in his article: Scientific Visualization: A Synthesis of Historical Data, gives an extraordinary compilation of historical examples that state the importance of visualization within the realm of scientific investigation. He collects descriptions of Archimedes, Edison, Newton, Tesla, Kekulé, Einstein, Faraday, Hawking, and Galileo.

Scientific thinking

Scientific thinking refers to the process by which scientists come up with reliable knowledge about nature or the universe. Often considered in parallel to critical thinking and analytical thinking, (Schafersman, Steven. (1997), scientific thinking relies on the use and application of the scientific method.

"The scientific method is practiced within a context of scientific thinking, and scientific thinking is based on three things: using empirical evidence (empiricism), practicing logical reasoning (rationalism), and possessing a skeptical attitude (skepticism) about presumed knowledge that leads to self questioning, holding tentative conclusions, and being undogmatic (willingness to change one's beliefs)". (Schafersman, Steven. (1997).

The interest on science in this investigation is focused on the relationship of scientists and the use of visual images to acquire knowledge, rather than the implementation of the scientific method. A close relationship between natural sciences and the use of visualizations appeared to be due to the objective of this area of science to create models of nature. (Hyrkäs, Johanna. (2000).

What is remarkable about the scientific thinking process is its effectiveness in the production of results. Observation of nature and experiments are ways of testing hypotheses and creating models that reproduce the original conditions in which a phenomenon was observed. Scientific thinking is the outcome of a series of processes that use empirical evidence. In scientific thinking and the scientific method, the steps are linear and have to follow a predetermined set of rules.

Scientists very often rely on observations and visual representation when they practice scientific or critical thinking, but the processes by which the results are achieved are not a reflection of a visual thinking process, the process that leads to scientific thinking is concrete in nature.

Design thinking

Design thinking focuses on finding solutions for problems of different nature through the implementation of creative processes. As broad as the concept of design is, design thinking applies to all the areas that practice design as a problem solving method, from architecture to graphic design.

Design thinking analyzes in depth the various aspects of the problems that it is asked to solve. A good way to understand design thinking processes is to understand the stages involved in its implementation. Stated in the article about design thinking in Wikipedia, there are seven stages: define, research, ideate, prototype, choose, implement, and learn. Each of these stages carries complex tasks that better explain some of the procedures involved in design thinking. Other important statement of this article is that these stages do not necessarily occur simultaneously.

Although the definitions or ideas about design thinking are still unclear and undefined, this fact does not impede us from finding relationships and differences between scientific thinking and design thinking. The comparison of these areas leads us to think in visualizations as a common element, and probably this exercise will help us to understand better the problems encountered with the implementation of each other's ways of thinking.

As opposed to scientific thinking, design thinking is more abstract and does not rely only on empirical evidence. Visual images are the core constituents of production in design thinking and very often designers use them as a model to represent findings and to exteriorize ideas.

Creativity in science and design

A careful analysis of science and design reveals that creativity is used equally in both fields. In reality, scientists use creativity to create new hypotheses that can be tested later in order to find verifiable facts, or to produce experiment designs that accommodate needs for specific solutions. Explanations of phenomena based on data collected and analyzed by researchers, also need creative processes that some times lead to the proposal of new theories.

The use of creativity in design is more global; it appears in every stage of the design process. The way designers confront their challenges implies finding creative solutions while defining the problem, researching, ideating solutions, prototyping, choosing the correct solutions, and implementing them.

Finding interconnections between creative processes in science and design allow us to highlight were these areas touch each other. Common grounds open the possibility for interaction and collaboration.

Understanding each other's fields

To understand a field from the perspective of those who know it and work on it demands attitude and determination. The better way to comprehend the connection between fields is to experiment and experience them by our selves. Conversations and discussions marked the beginning of the interaction between the scientist and the designer involved in this research. It is also important our commitment in producing effective ways of communicating. For these reasons, scientists and designers need to share knowledge and experiences in order to acquire the information necessary in the creation of successful visualizations that effectively communicate scientific content.

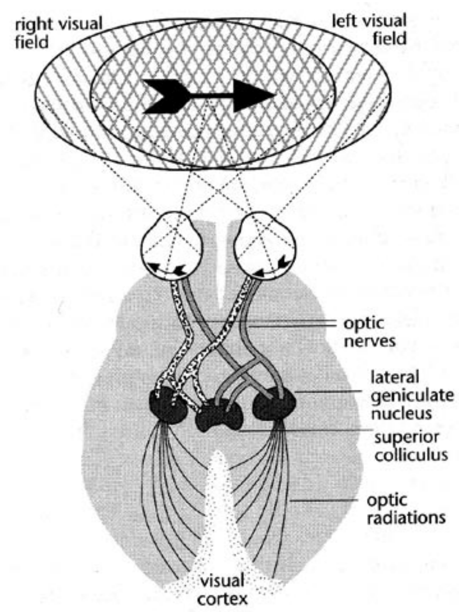
"Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency. Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space".

Edward Tufte

5 / SOME PRINCIPLES

principles of design

The following is an annotated list of the most important principles and concepts used in the development of the prototyped interfaces. Must of these principles were derived from Colin Ware's book "Information Visualization: Perception for Design". This book compiles an excellent collection of theories and studies about the design and visualization of data.



The physiology of perception.

Image from: *Alan MacEachren's How Maps Work*.

Visual perception

Most of the principles used in these tools are based on their application for perception and the influence that the result of using them has in perception processes. The success of visualization is usually dependent on our abilities to perceive, process, attend to, and memorize the elements present in that visualization.

Color

The use of color in screen interfaces has many aspects and innumerable applications, for the purpose of the interfaces developed in this thesis; color has a very specific intention; which is to assign colors to important elements in the interface so these elements are distinguished from the rest. In the Sequence Generator, color is assigned to the individual parts of the body touched and to the sequences conformed by three of these parts.

In the case of shapes, the use of color is even simpler since there is practically one full contrast between the background of the interface and the shapes. Other aspects of color in shapes are the low contrast between the background and the menus and the consistency of the three colors used.

Being able to establish relationships between specific colors and the parts of the body ease the process of perceiving and recognizing those parts, as well as applying methods of organization to the sequences.

As stated by Colin Ware in his book, "...as with lightness contrast, the ultimate purpose of the contrast-causing mechanism is to help us see surface colors accurately by revealing differences between colored patches and background regions".

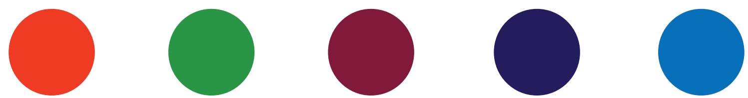
The colors used in the tools were the result of studies between both scientist and designer. One from the perspective of the person who will be interacting with the tool and the other to assure that the chosen colors would comply with general rules of perception and interface interaction.

"Color can be extremely effective as a nominal code. When we wish to make it easy for someone to classify visual objects into separate categories, giving the objects distinctive colors is often the best solution". (Ware, 2004)

Because the elements to be labeled with color coding were a small set, contrast with background was not an issue. Colin Ware also mentions that applying a thin white or black border to the color-coded element is a method for reducing contrast effects between backgrounds and color-coded elements.

Colors used for color-coding in the Sequence Generator match some in the list of recommended colors for color-coding from Colin Ware's book.

If future development of the tools is in mind, there will have to be more investigation in aspects such as colors assigned to new elements automatically by the tool or chosen by the user.



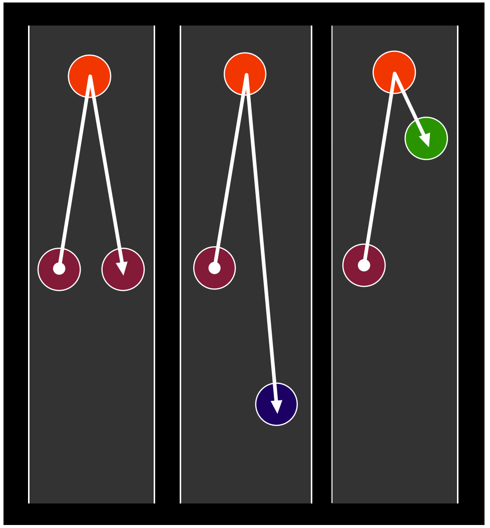
Attention

Because the users of specific interfaces and interactive tools intended for data mining are in close interaction with large amounts of data, visualizations have to display visual elements with special care on attention. Aspects such as proximity between elements and similarity are key to assure a successful interaction.

In the case of the interfaces in the *Mapping Science* Lab, large amounts of data are not an issue, but taking in consideration basic principles of composition always help to the efficiency of the tools. To be more specific, the Sequence Generator compiles a series of sequences that are displayed in horizontal order, allowing the user to compare and analyze the way these sequences appear grouped. This problem was resolved by making the sequences individual elements that can be distinguished from adjacent similar sequences.

Both tools seek to be experimental in terms of acquiring familiarity with their use. The use of motion in interactive tools is common when attracting attention for certain graphical elements. Ware, (2004). In the case of the Sequence Generator and Shapes, this is not necessarily true since some times motion can distract users from the activities that they are performing.

Application of the principles and theories from Ware's book in the realm of motion are targeted to the detection of elements displayed in the interactive tools. As Ware describes in his book, "A concept called the useful field of view has been developed to define the size of the region from which we can rapidly take in information", and that the useful field of view can be improved by the use of motion. So the display of data in the two case studies only uses motion when new elements appeared in the interface, to direct users attention to those new elements so new data does not goes unnoticed.



Sequences showing the use of color, contrast, and other issues such as adjacency and proximity.

Image from: *Mapping Science. The Sequence Generator.*

Pre-attentive processing

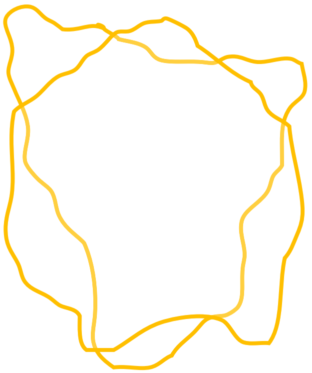
"We can do certain things to symbols to make it more much likely that they will be visually identified even after very brief exposure. Certain simple shapes or colors "pop out" from their surroundings. The theoretical mechanism underlying pop-out is called preattentive processing because logically it must occur prior to conscious attention. In essence, preattentive processing determines what visual objects are offered up to our attention. An understanding of what is processed preattentively is probably the most important contribution that vision science can make to data visualization".

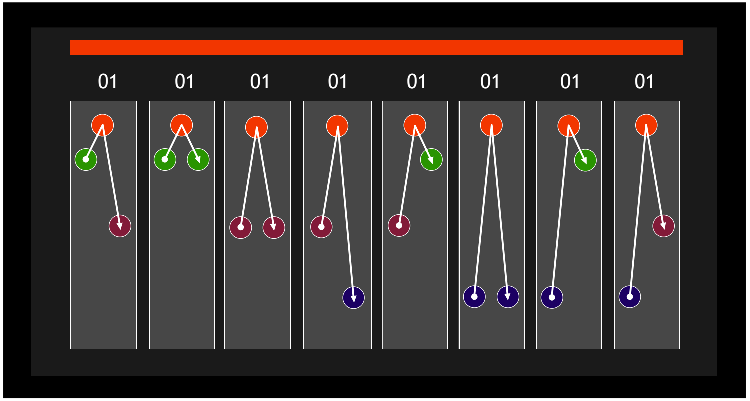
Basic application of the rules of pre-attentive processing in the Sequence Generator and Shapes were treated by the aspects mentioned above, (i.e., color and motion).

Other concepts such as rapid area judgment and conjunction with spatial dimensions are also used but in less intensity. The rapid area judgment was studied for measurement and judgment of areas of the shapes, and conjunction with spatial dimensions applies to the distribution of sequences in the *Sequence Generator*. To better understand the Conjunction with spatial dimensions lets quote another explanation by Ware. "Searches can be preattentive when there is a conjunction of spatially coded information and a second attribute, such as color or shape." This principle is illustrated by the design of the sequences in the *Sequence Generator*.

When designing features like the comparison tool in Shapes, concepts like rapid area judgment were important.

Image from: *Mapping Science. Shapes*.



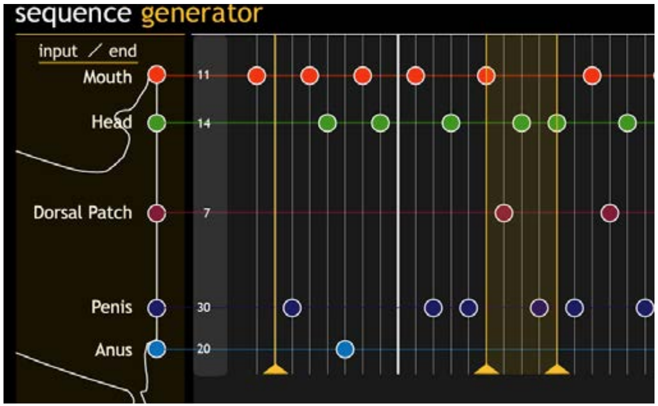


Relationship between background and elements is shown here. Individual texture for each sequence makes the interpretation easier and separates the elements perceptually.

Image from: *Mapping Science. Sequence Generator.*

Highlighting

Another important concept in the design of interactive interfaces for data visualization is that of highlighting. When simple objects need to stand out from the rest of the data, many techniques of highlighting become useful. Highlighting was used in the *Sequence Generator* to solve a simple problem when lines in the input area had to be selected and then a marker needed to be applied.



Highlight in the Sequence Generator, differences between selected and unselected features are highly benefited by highlighting.

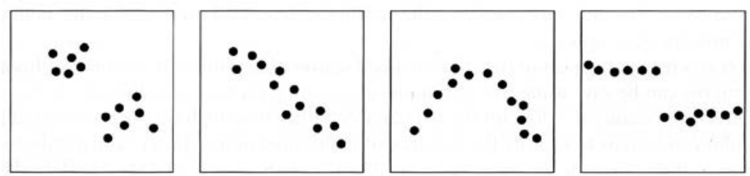
Image from: *Mapping Science. The Sequence Generator.*

About patterns in visualization

" Understanding pattern perception provides abstract design rules that can tell us much about how we should organize data so that important structures will be perceived. If we can map information structures to readily perceived patterns, then those structures will be more easily interpreted." Ware, (2004).

" Learning is important in the pattern mechanism. It occurs in the short term through visual priming and in the long term as a kind of skill learning. Priming refers to the fact that once we have seen a pattern, it becomes much easier to identify on subsequent appearance. Long-term learning of patterns occurs over hundreds or thousands of trials, but some patterns are much easier to learn than others (Fine and Jacobs, 2002)." Ware, (2004).

These quotes are intended to explain the importance of patterns and pattern recognition in data visualization. As per my investigation, I consider pattern recognition as the primary component of both, creative insight in producing hypotheses for science and the process of interacting with data through visual thinking.



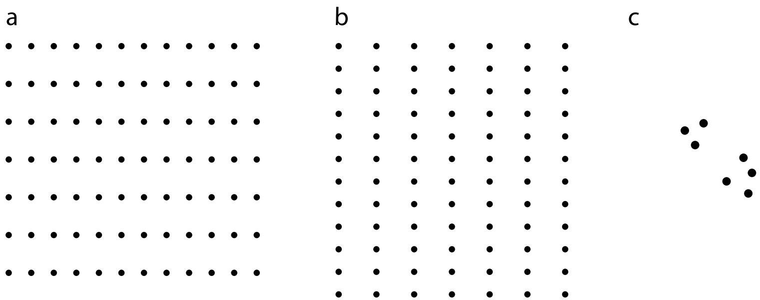
Simple dots show how graphical elements display patterns.

Image from: Colin Ware's: Information Visualization.

Gestalt laws

One of the first if not the first attempt to conceptualize about patterns in visualization is that of Gestalt Laws. These laws provide an extensive study of how patterns are perceived and processed. Many of these laws were considered for the design of the prototypes in Mapping Science and will be shown here by reproducing the examples in Ware's book.

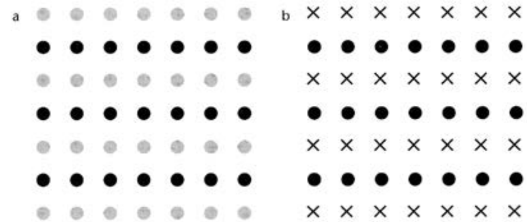
Proximity



"Spatial proximity is a powerful cue for perceptual organization. A matrix of dots is perceived as rows on the left (a) and columns on the right (b). In (c), because of proximity relationships, we perceive two groupings of dots".

From: *Colin Ware's Information Visualization.*

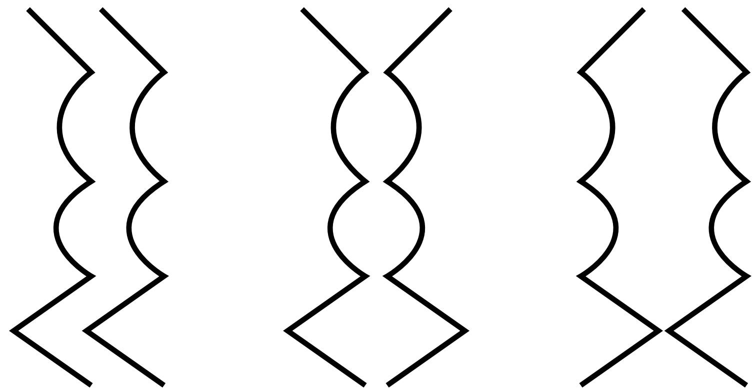
Similarity



"According to the Gestalt psychologists, similarity between the elements in alternate rows causes the row percept to dominate".

From: *Colin Ware's Information Visualization.*

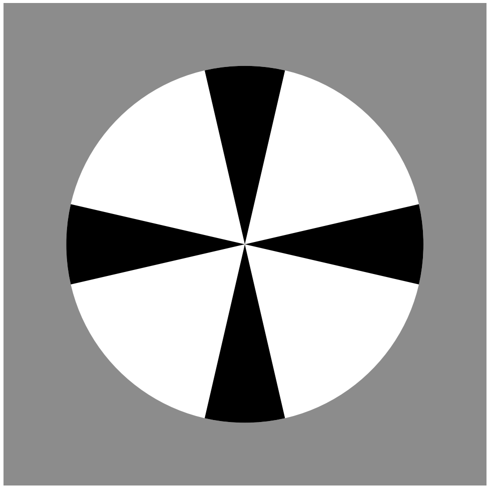
Symmetry



"The pattern on the left consists of two identical parallel contours. In each of the other two patterns one of the contours has been reflected about a vertical axis, producing bilateral symmetry. The result is a much stronger sense of a holistic figure".

From: *Colin Ware's Information Visualization.*

Relative size



"In general, smaller components of a pattern tend to perceived as objects. In the figure, a black propeller is seen on a white background, as opposed to the white areas being perceived as objects".

From: *Colin Ware's Information Visualization.*

The Gestalt laws listed here were those used in this thesis. My personal recommendation for those interested in specific applications of the laws in data mapping and visualization is: (Colin Ware's Information Visualization) and (Alan MacEachren's How Maps Work).

These laws, rules, and principles were of invaluable help for the production of my prototypes and will, I hope, be an aid in understanding my approach to design solutions applied to the specific experiment of my thesis.

6 / ABOUT DATA AND INFORMATION

data & information

Defining data is often closely related to specific needs of the field in which data is contextualized. General concepts can be used and applied to almost anything, and specific definitions usually refer to very precise elements. There are also other considerations when defining data such as if data is the principal component of information, what becomes of data when is organized, the abstract nature of data, and so on.

To better see how broad and varied the definitions of data are, I present a selection that I considered the most representative, taken from the article "Conceptual Approaches for Defining Data, Information, and Knowledge" Zins (2007).

> "Datum is every thing or every unit that could increase the human knowledge or could allow to enlarge our scientific, theoretical or practical knowledge, and that can be recorded, on whichever support, or orally handed. Data can arouse information and knowledge in our minds". (Maria Teresa Biagetti)

> "Data are the basic individual items of numeric or other information, garnered through observation; but in themselves, without context, they are devoid of information". (Quentin L. Burrell)

> "According to Stonier (1993, 1997), data is a series of disconnected facts and observations. These may be converted to information by analyzing, cross-referring, selecting, sorting, summarizing, or in some way organizing the data". (Gordana Dodig-Crnkovi)

> "Data typically means the "raw" material obtained from observation...". (Hamid Ekbia)

> "Data are facts and statistics that can be quantified, measured, counted, and stored". (Donald Harmon)

> "Data are atomic facts, basic elements of "truth", without interpretation or greater context. It is related to things we sense." (Donald Kraft)

> "Data are commonly seen as simple, isolated facts, though products of intellectual activity in their rough shape". (Jo Link-Pezet)

The list of definitions is long, but these few allow us to reach a better understanding of data and its multiple applications. By analyzing these definitions we could conclude that: (1) there is an obvious relationship between data and information, (2) in almost all the definitions, data is a raw or individual element, or a collection of individual elements, (3) data is a product of either observation or perception, (4) data does not have context, (5) data can be gathered, stored, and organized, and (6) data is the product of intellectual activity and its analysis demands awareness.

Communicating a message through visualization is mostly achieved by visual representations of data, information, or knowledge. These can be considered the basic components of messages and communication processes. For this reason it is equally important to define information. Lets analyze definitions of information extracted from the same article "Conceptual Approaches for Defining Data, Information, and Knowledge" Zins (2007).

"Information is data that has been processed into a form that is meaningful to the recipient". (Davis & Olson, 1985)

"Information is that which is conveyed, and possibly amenable to analysis and interpretation, through data and the context in which the data are assembled". (Quentin L. Burrell)

"Information is the sum of the data related to an entity". (Henri Jean-Marie Dow)

"Information is organized data (answering the following basic questions: What? Who? When? Where?)". (Nicolae Dragulanescu)

"Information. Collocations of data that become meaningful to human beings—e.g., as otherwise opaque units of binary code are collected and processed into numbers, artificial and natural languages, graphic objects that convey significance and meaning, etc". (Charles Ess)

"Information is data that has been categorized, counted, and thus given meaning, relevance, or purpose". (Donald Hawkins)

"Information is facts, figures, and other forms of meaningful representations that when encountered by or presented to a human being are used to enhance his/her understanding of a subject or related topics". (Haidar Moukdad)

By reviewing these concepts we can imply that: (1) Information is related to data but not limited to data, (2) Information always involves meaning and context, (3) Information is the result of human interaction, communication, and experience, (4) analysis and interpretation are important in the process of understanding information, and (5) as opposite to data, information is organized and structured.

To understand data is to know the connections between each single part of a whole. A datum is like the neuron of information and information is like a neuronal network where connection between its elements is essential for the sake of the final result, whether it is a synapse or a piece of information being conveyed.

The following definitions are the result of my own investigation and analysis of existing materials and they may not apply to a broad application of the concepts of data and information.

Data is the result of gathering and storing individual items or datum, which are the outcomes of human's interaction with the environment. Data can be measured, interpreted, represented, and studied, but has no meaning, structure, or context by itself.

Information is the product of organizing and processing data or any other item or group of items that result from human experience and humankind's intent to communicate.

The importance of defining data and information relies on the fact that they are the core elements to be visualized when a message is being conveyed through visual representations. Visualization of information, responds to the primary goal of communicating, while visualization of data, can be used in manipulation, interaction, storing, and representing of that data so these processes can be more effective and communicate better the content carried by the data.

The data used in this investigation has unique characteristics. It is quantitative in nature, is represented by numbers, but more importantly, is the product of external factors related to the presence of the dorsal patch. Numbers are in all cases the result of analysis and interpretation of the factors, but they represent different types of data. For example, numbers can represent relationships between behaviors, repetitions and frequency of a specific behavior, values, measurements, dates, amount of parasites, comparisons, etc. In some cases, numbers are the result of an interpretation of qualitative aspects that need to be measured in certain ways. In the case of behavior, many factors that are in essence qualitative are defined and quantified. The final outcomes are numbers but the main types of data differ in their nature.

To be more specific, both case studies of my thesis use qualitative and quantitative data. In case study # 1 (The *Sequence Generator*), sequences, are more qualitative, and repetitions, are more quantitative. In case study # 2 (*Shapes*), the main focus is in morphological traits (qualitative) of the dorsal patch and the connections between these traits and other data such as number of parasites, dates of collection, and measurements (quantitative). More information and details about the uses and findings of the data will be given later in this document when describing each case study individually.

MOUTH		
dorsal patch - mouth - penis	1	0
dorsal patch - mouth - head	8	1
dorsal patch - mouth - dorsal patch	0	3
head - mouth - head	4	0
head - mouth - dorsal patch	2	4
penis - mouth - head	17	2
penis - mouth - penis	1	0

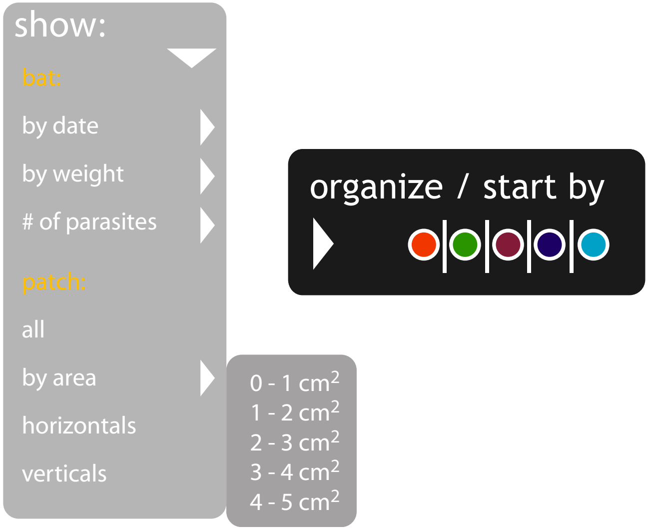
A segment of the original table shows how the scientist used to present the data related to the sequences.

From: *Mapping Science*.

Organizing data

Another main issue when dealing with visualization of data is its organization. Complex content is a primary concern at the moment of organizing data together, and how the content is conveyed depends on how we decide to organize the data. Richard Saul Wurman, in his book Information Anxiety clearly states that the ways of organizing data are finite, location, alphabet, time, category, and hierarchy give name to his LATCH principle. Different ways of organizing data allow us to highlight different attributes of that data.

Another important part of my investigation seeks to find effective ways to organize data and lay them out on a single interface, so the interaction between the scientist and his/her data is the most productive one. Improving the ways a scientist interacts with data (e.g. input, analyze, and visualize) depends not only on translating data into a visual form but on organizing the data so the interaction gets benefited by this organization.



Menus represent ways of displaying data that have been previously organized. Sometimes filters can be the window to organize and structure data.

From: *Mapping Science. Sequence Generator and Shapes.*

“The most important instrument of thought is the eye”.

Benoit Mandelbrot

7 / WHY BATS ?

why bats?

Bats are of primary interest in natural sciences because they are: (1) the only group of flying mammals, (2) the most successful group after the rodents in terms of diversity and distribution, (3) one of the few mammals that have developed sophisticated ways of communication (i.e., auditory and olfactory signals), (4) key to control insect populations and pests, and to maintain the equilibrium of forests by dispersing seeds and pollinating nocturnal plants.

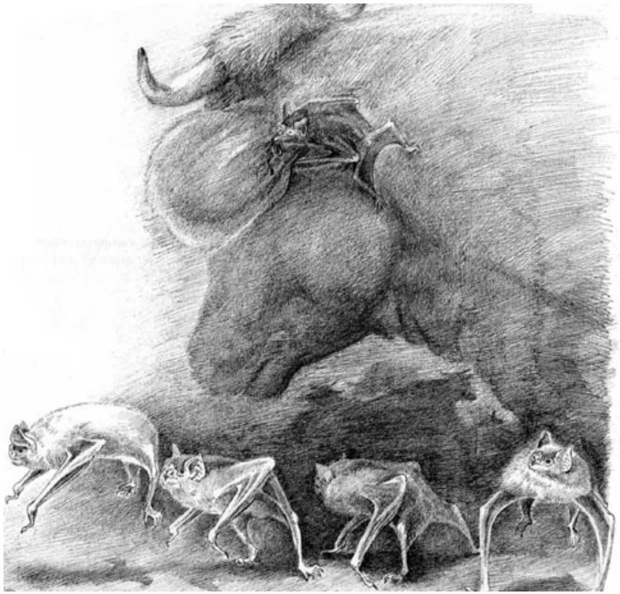
Bats represent 25% of the total species of mammals of the world and are found in almost every environment on the planet. Their ways of communication have inspired several fields of study such as engineering, medicine, and physics. As the result of research performed on bats, engineering has developed radars and detection systems, medicine is allowing hearing-impaired people to explore space and become part of the society, and physics has studied wave propagation and ultrasound.

There are about 1100 species of bats in the world; all of them show different evidences of adaptation, from feeding to roosting habits. Depending on what their feeding habits are, bat's bodies display differences in physical and morphological features, for example, a large uropatagium (i.e., membrane attached to legs and tail) is used to capture insects whereas bats that feed on flowers do not have uropatagium at all.

This bat is using the tail membrane to capture an insect. This membrane is an example of the great adaptation to the environment that bats have developed.

From: *Klaus Richarz & Alfred Limbrunner. The world of Bats.*

Roosting habits also originate morphological changes in bats. There are bats that live in slippery banana tree leaves and have developed suction pads in their wrists allowing them to hang from the leaves without falling down. The common vampire bat sucks its victim's blood and returns to its roost crawling, which is why it has developed a hard pad in its thumb to help him to crawl.

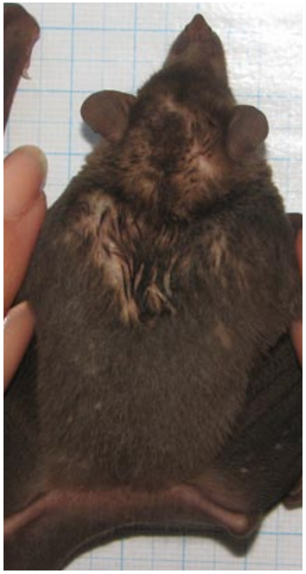


This image shows the incredible behavioral adaptation of the vampire bat.

From: *Klaus Richarz & Alfred Limbrunner. The world of Bats.*

The Dorsal Patch

The dorsal patch is one perfect example of the adaptations mentioned above as it has behavioral, chemical, communicational, and evolutionary connotations. It also brings a novel phenomenon to the field of bat research, as it is thought that the patch could have implications in sexual selection, protection against parasites, communication, and attraction of females. The fact that is a feature present only in males makes it of interest and the patch could be a key player in the future and success of this species.

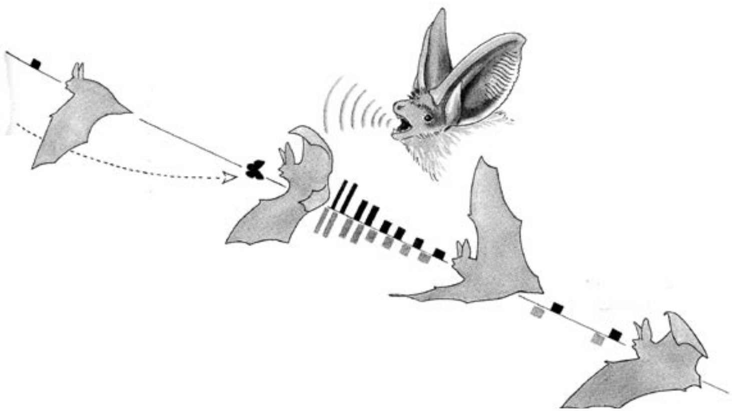


The dorsal patch as an evidence of adaptation.

Image from: *Muñoz-Romo.*

Olfactory communication in nocturnal mammals

Different needs demand different solutions in nature, and species of mammals around the world have developed modes of communication that fulfill their necessities. Elephants use infrasound and their strong sense of olfaction to detect and communicate with other members of their herd and to recognize one of their own that has deceased by smelling its bones. Wales and dolphins use ultrasound instead to communicate.



Echolocation.

From: *Klaus Richarz & Alfred Limbrunner. The world of Bats.*

As nocturnal mammals, bats have developed different nighttime communication methods, because at night, their eyes are not much help. Olfactory and acoustical cues become crucial when light is deficient, and some bats had perfected echolocation, and chemical signals as communicational modes.

Contrary to the common belief, bats are not blind, it is just that their sense of vision is not as well developed as their other senses. Bats are active at night when they have to rely on acoustical signals or vocalizations that bounce on objects and are perceived by bats' strong sense of hearing on their way back, a phenomenon known as echolocation. Echolocation is perfect for hunting and locating objects that block bats' flight, however, in the constrains of a cave, between thousand of other immobile bats, it might be necessary to complement this method with others like olfaction. Indeed, olfactory signals are paramount to the processing of information about conspecifics, family members, or potential mates.

Chemical communication

Diverse adaptations have been produced to complement olfactory biological processes and bats use chemical signals in order to communicate information such as health, reproductive condition, resistance against parasites, and individual quality. These chemical signals are very effective as a communication tool because their interaction with olfactory receptors and nerve endings convey the information accurately and rapidly. Consequently, inefficient communicational strategies are substituted by chemical signals, which are more suited for the low light and crowded conditions of the environment in which the bats inhabit.

8 / ABOUT THE SCIENTIFIC INVESTIGATION

the scientific investigation

As stated in the title "The dorsal patch and chemically mediated courtship behavior of the long-nosed bat, *Leptonicteris curasoae* in Venezuela", the scientific research that serves as experimental ground for my thesis, explores the ecological and potential evolutionary significance of a structure found in bats.

As background, the following definitions are presented.

> Ecology is: "The study of interactions of organisms with one another and with their physical environment". Peter Raven et. al., (2005).

> Evolution is: "Genetic change in a population of organisms; in general, evolution leads to progressive change from simple to complex". Peter Raven et. al., (2005).

The patches are only present in males; this can establish a precedent to connect this sexual dimorphism (i.e., differences between sexes of the same species), with reproductive aspects of the studied species.

The study of the dorsal patch in this bat species is very important in terms of its ecological and evolutionary significance. Munoz-Romo discovered that dorsal patches have a diverse list of chemical compounds, some of which could be functioning as protection against ectoparasites.

The dorsal patch in one of its early stages.

From: *Muñoz-Romo (Ph. D. Thesis)*

In this screen shot two bats are showing interaction related to the patch. Smells and other traits of the patch could be functioning as communication cues.

From: *Muñoz-Romo (Ph. D. Thesis)*

On the other hand, a specific chemical or a combination of chemicals could be also functioning as the 'identity' for each male. These results would be indicating that those males that develop dorsal patches would have advantages over those who do not. Females would prefer males with fewer parasites, and also it would be easier and quicker for the female to find and recognize a male that has a chemical identity that facilitates encounters.

It has also been demonstrated that odors from males are reflecting their health condition, so that females can select healthy individuals to mate. It is also important to consider that males develop dorsal patches exclusively during the mating season, which suggests the importance of dorsal patches during reproduction and then for the future of this species of Neotropical migratory bat.

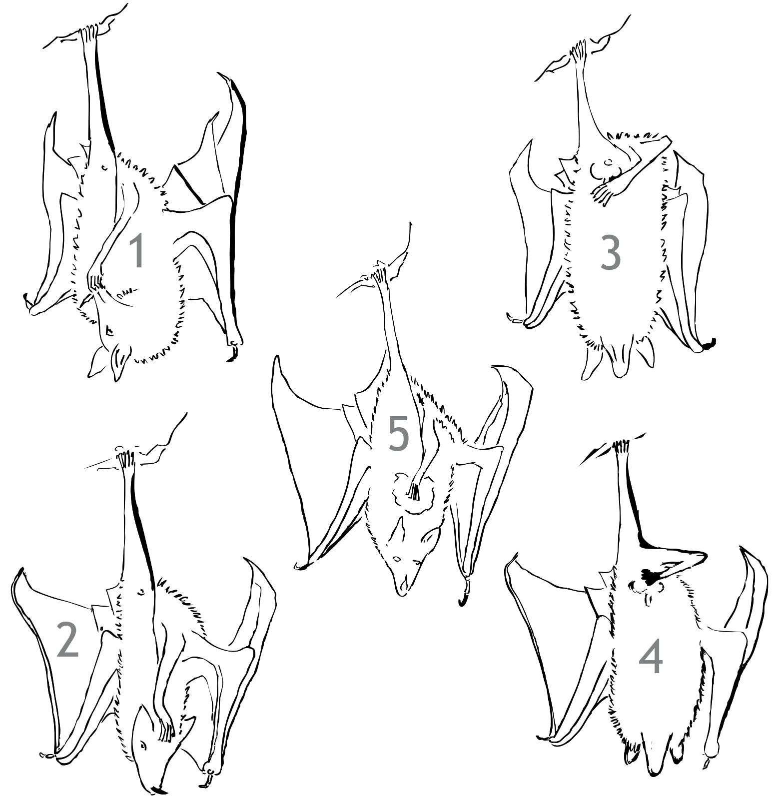
Not all the parts of the research project can be translated into the visual domain. Many aspects seem to be specific to the scientific realm, while others are more suited for translations and implementation of images that help to explain the content carried by them. For example, the ecological and evolutionary relevance of this study relies in part on specialized scientific knowledge and experiments that help to find answers to specific questions such as, is this dorsal patch phenomenon related to physiological changes in the organism? When the dorsal patch first appeared in the evolutionary history of the long-nosed bat? What physiological profile is necessary for the development of a dorsal patch? The answers to these questions demand an analytical process that is more characteristic of the scientific thinking.

On a different level, we encounter other kinds of information and data that are much more appropriate for the use of images, diagrams, and graphs. Measurable and qualitative data are easily represented visually and allows the representations to reveal hidden features of the data.

Two primary elements of the research project were key to both the scientific investigation itself and the application of visual solutions to the needs of this investigation. These two elements were, the study of behavior related to the dorsal patch and morphological traits of the patch with a deep focus on symmetry.

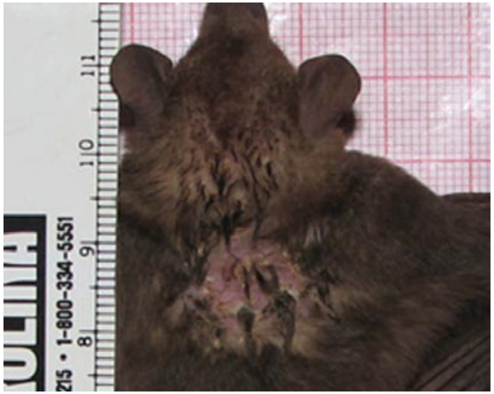
Behavior

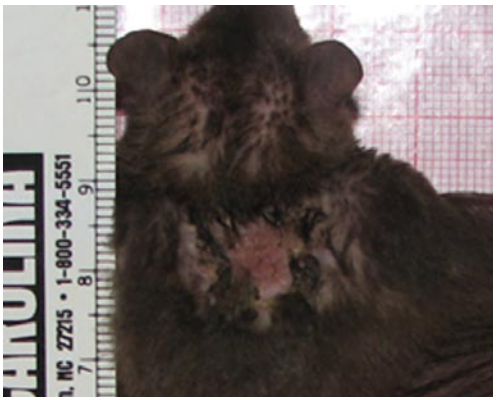
The dorsal patch is apparently the result of a complex behavior. The study of this behavior through observation and posterior analysis of the outcomes of these observations represented an opportunity to improve the interaction between the scientist and her data. By measuring behavior and studying the relationship of it with the production of the patch, the scientist defines a variety of rules that help her to comprehend other connections between evolutionary and ecological aspects of the patch.



Four parts of the body to take fluids from, finally smearing the fluids in the dorsal patch conform the basic behavior studied by the scientific research.

From: *Mapping Science*.





The characterization of the varied shapes for the patches is important to establish relationships between aspects like morphology and physiology of the bats, which are decisive for the research.

From: *Mapping Science*.

This project includes behavioral observations because these usually provide answers for questions related to the potential function of structures new in science such as the dorsal patch. As a result of investigating behavior, the scientist discovered that the patch was not the result of glandular activity, as suspected before, but the outcome of a complex behavior performed by male bats.

Shape

The study of the shape of a dorsal patch is fundamental to establish a generalization that allows the researcher to describe the dorsal patch as a particular shape that characterizes it. The characterization of shape for each one of 69 dorsal patches included in the study is a difficult task given that each dorsal patch shape needed to be individually described (i.e., delimitation and estimation of area) and subsequently compared. The visual comparison would allow the researcher to see whether a specific shape dominates within the sample. Ejemplo aqui.

The researcher answers questions such as: Is there a pattern or trend in dorsal patch shape? How difficult is to discover this pattern? Dorsal patches have remarkable differences between them, and shape seems to be an important factor that needs to be studied, despite the difficulty of investigating a flexible trait.

Symmetry

Symmetry of forms is a very important topic in biology because of its relevance in mate choice and sexual selection. It has been demonstrated that symmetry in a bilateral body and its components has a strong relationship with an organism's physiological equilibrium.

Thus, the more symmetric the individual (i.e., similarity of left and right sides), the more physiologically equilibrated it would be. A physiological equilibrated individual is better suited to perform the basic activities (find food, defend territories, attract conspecifics, and reproduce) and would be more successful in terms of reproduction.

Comparison of dorsal patch shapes to study symmetry is a challenge because the dorsal patch is a flexible trait. Flexible traits are very difficult to measure because they depend on the elements interacting with it (i.e., researcher measuring the trait) as opposing to a non-flexible feature (i.e., bone).

In the particular case of the dorsal patch of *L. curasoae*, it is hypothesized that high symmetry in dorsal patches would be related to chemical signals produced by healthy males. Asymmetrical dorsal patches would be emitting a chemical signal that would indicate a handicap (e.g., illness) of the sender.

These three elements, which indeed became two, (since symmetry falls into the study of shapes and morphological characteristics of the patches), represented the common ground for visualizations and development of tools to be applied to the scientific investigation in bats. The scientific data derived from these areas of the investigation not only was suited for the task of visualizing it, but was remarkably interesting as well for experimentation that allowed the scientist to start interacting with her data using visual thinking processes.

9 / THE CASE STUDIES

case studies, an introduction

Extensive conversations with the scientist in charge of the research on bats, which happens to be my wife, made me realize how different scientists perceive concepts that pertain to the field of design, and how they separate the tasks undertaken by designers and scientists.

My first reaction was to convince my wife through argumentation that design was much more than what she thought. This was not going to be an easy task and, after fruitless discussions, I realized there was a mayor impediment underneath that took me sometime to appreciate. The incapacity of agreement and understanding from both sides in some of the points discussed went beyond my reasoning. Obviously the way our brains work was different and our professional training took over our ability to reason.

A good solution was to make this discussion more practical than theoretical. I tried to explain my points by illustrating them with visual images that made them clearer. Eureka! Arguments started to change and minds as well, from that point on everything was easier.

Getting into an understanding led both of us to be able to assess each other's field from a different perspective. Discussions of this sort were held with other scientists and the results were always the same with varied degrees of acceptance or rejection of ideas or arguments. In most of the cases the problem was a lack of understanding as opposed to indisposition, which was always solved by using visual language to illustrate the points.

All the interaction produced from sharing ideas and information about our fields of knowledge originated a deep interest to exteriorize and analyze this experience and my wife's thesis was a perfect opportunity. I started by studying some aspects of science so I could understand basic concepts used in her research project, followed by simple applications of my studies in visualization to improve data analysis and representation.

First, applications of visualization were intended to be an aid to the thesis on bats but they quickly became more complex, allowing me to investigate and study the components of this interaction and produce my thesis at the same time.

Collaborations took more than eight months and provide me with the raw material that I needed to make strong case studies that could explain my ideas. Primarily focusing on the interaction between scientist and designer as important parts of the creation of effective visualizations that carry complex scientific content This exercise provides the ground for scientists to experience visual representations of their data that allow them to analyze it from a different point of view.

As a result of interactions and translation of the raw data into graphical elements, new needs became evident. For example, the process of interpreting and manipulating the data required simpler ways to accomplish it. Input, organization, and storing of data were the product of tedious processes and did not compile the results into a coherent collection.

One of the advantages of visualizing data is that of bringing the data into a more abstract language that still carries its definitions and allows that data to be represented and manipulated under common rules. Relationships can still be established between the quantitative nature of data and its visual representation. But it is the representation that facilitates accommodations to be applied to the data so it fits visual thinking processes.

Visual thinking processes are more intuitive and easier to process mentally, creating a suitable environment when patterns and characteristics of data need to be stored in memory and compared with other aspects. The purpose of this thesis is not to diminish the significance of scientific thinking processes but to highlight the importance and possible implementation of visual thinking processes in the development of scientific investigation.

The result of these interactions between scientist and designer, and between a scientific thesis and one in design produced two case studies. The *Sequence Generator* and *Shapes* are prototypes for interactive visualization tools to facilitate processes of interaction, interpretation, presentation, and communication of scientific content within the specific research on bats.

The environment of computers and their applications through the use of algorithms represent an excellent medium for the organization, storing, and structuring of data. In the same order, interfaces allow users to access the information in a single place and to perform several activities related to the data that will translate into efficiency in terms of time, effort, and outcomes. Worthy to mention is how advanced some computer applications are in visualizing data. In conclusion, dynamic media seems to be the most appropriate medium for the development of these tools and the one that offers more possibilities.

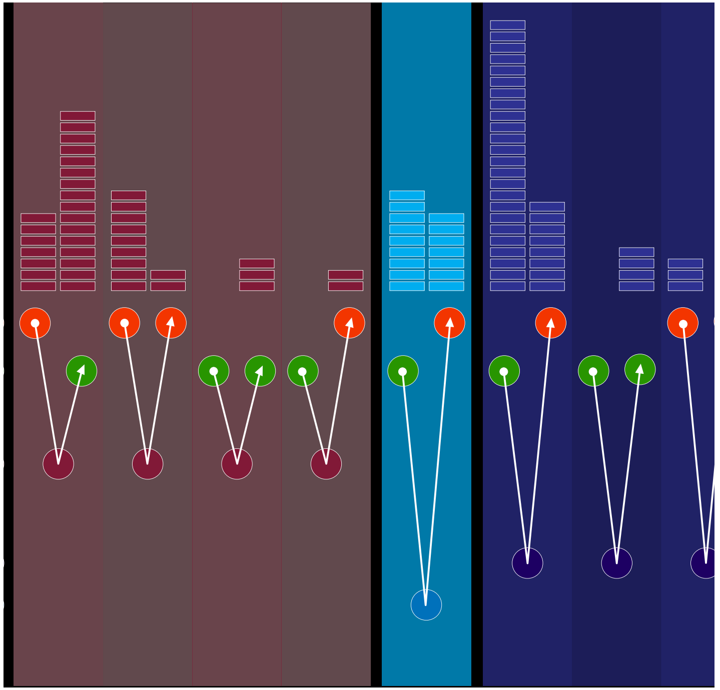
case study # 1 / the sequence generator

The *Sequence Generator* is a response to the needs for visual representation of behavioral aspects of the research on bats.

The first attempt to produce visual materials for this research, was the production of a diagram that translated the data represented in a table into visual elements. The table was in this case an appropriate visual solution for the representation of raw data, but not as effective when qualitative properties of the data were the focus of interest. Highlighting patterns within the data was hard and the result of exhaustive analysis was inefficient.

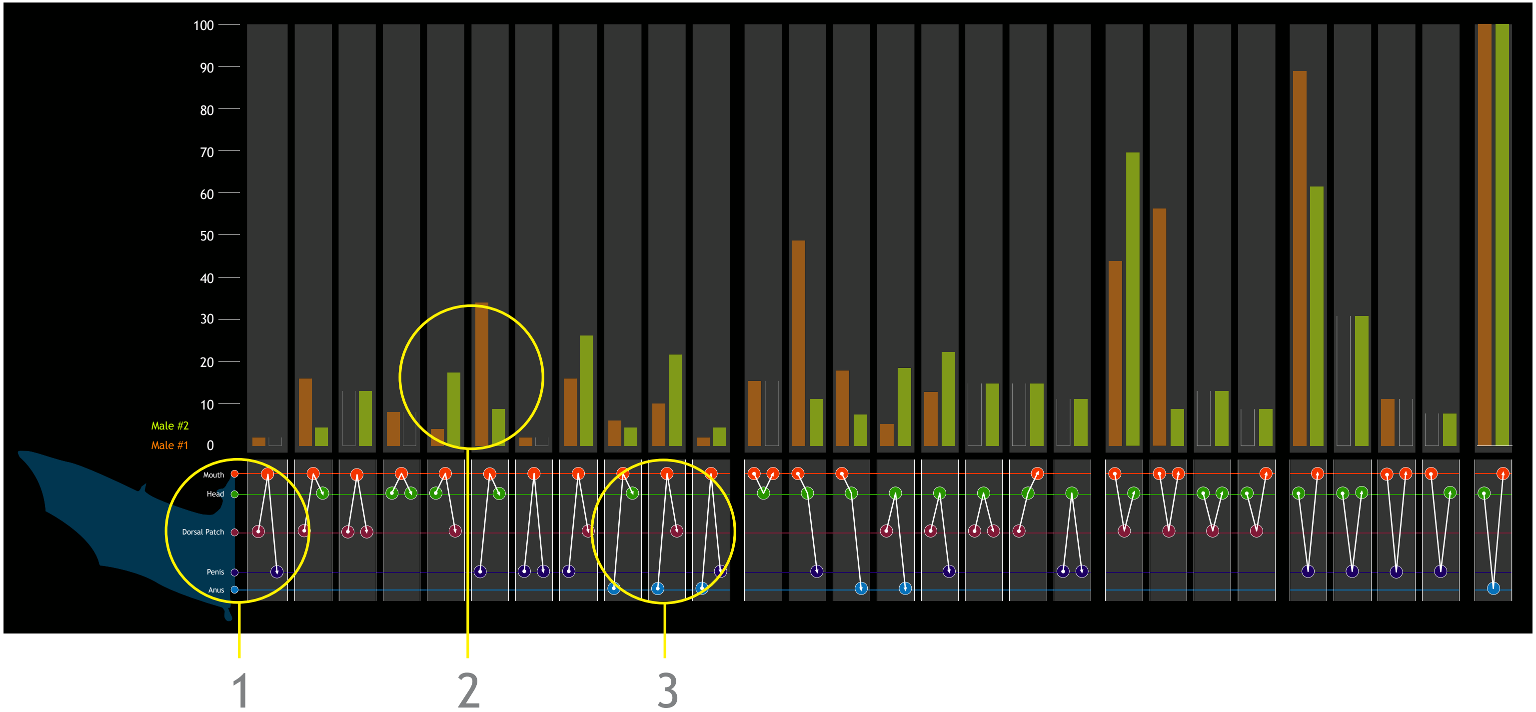
Visualizing data has many positive aspects but one of the most important ones is that it provides the ideal conditions for the data to be recognized and memorized with a minimum effort. This factor was key for the creation of the static diagram that translated the table's data.

Two behavioral elements were considered as primary, the need for visual representations focused on individual parts of the body touched by the bats and sequences created by studying groups of three parts touched. Basically, the sequences are conformed by a part of the body and that part touched before and after.



Third version of the static diagram that originated the design for the parts of the body and the sequences.

From: *Mapping Science*.



First diagram created with translation of original data to graphical elements. 1. Parts of the body and relationship with the bat. 2. comparison between two males, repetitions of behaviors. 3. Sequences conformed by three body parts.

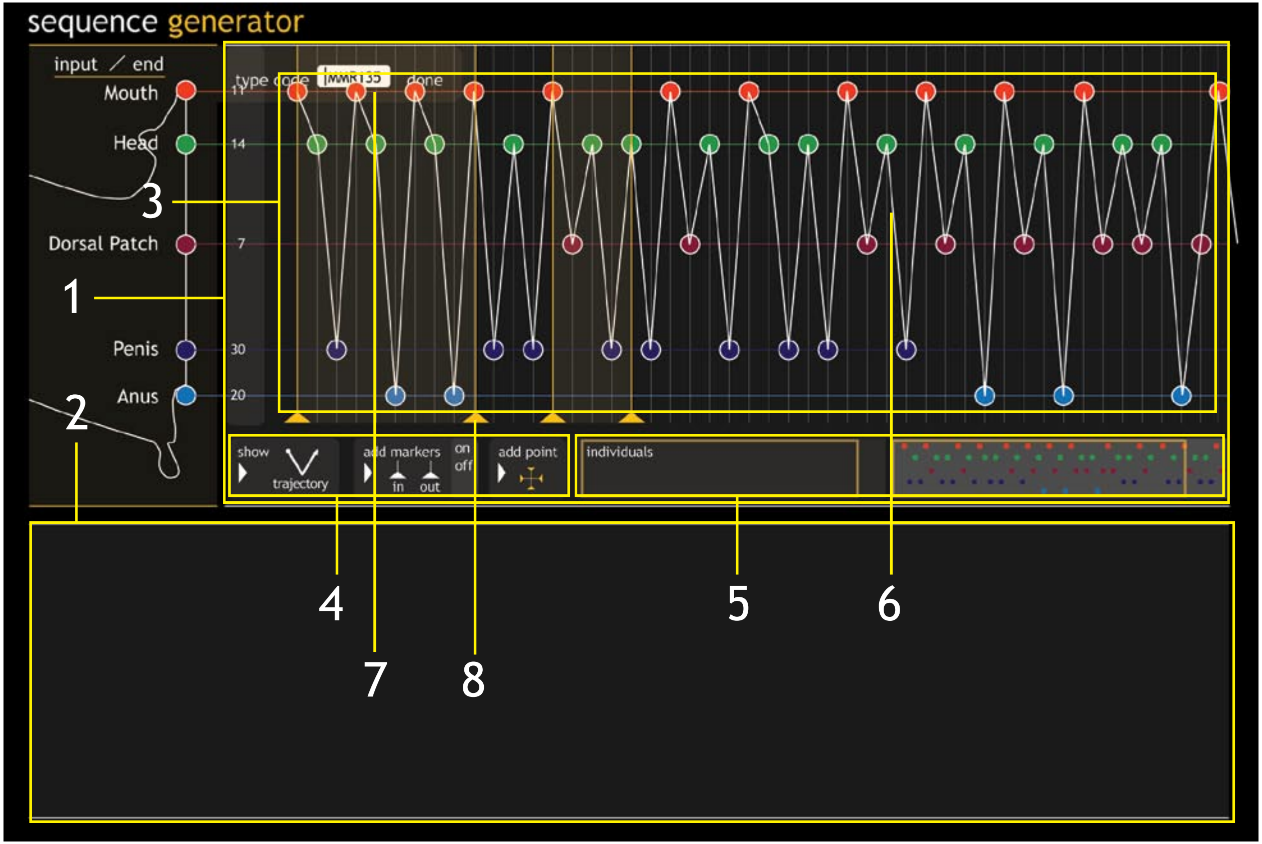
From: *Mapping science*.

The importance of the sequences relies on the relationship between these sequences and other aspects of the investigation, for example, the fluids that the bats smear on their backs carry chemical compounds that have different purposes in nature (e.g., parasite repellent or female attraction), these compounds are the result of combining body fluids in specific concentrations, which proves the connection between the compounds and the amount of times that a certain individual touches a specific part of the body; as well as what are the roles of each part of the body (e.g., it is presumed that the saliva coming from the mouth serves as a solvent for other thicker fluids).

Results of representing these elements visually also demanded other complex tasks such as the representation of relationships within the sequences, comparisons between two males, and repetitions of individual sequences. It became apparent to me that the scientist was lacking not only from an adequate representation of the data but also from a tool that facilitated the interaction with the data. The methods used to record and store the data were tedious and time consuming.

Working directly with a scientist in creating interactive tools to address their limitations with data representation and interaction has established a precedent of the kind of improvements that need to be accomplished in the field of scientific visualization.

The Sequence Generator is a screen based application and interface that seeks to improve the possibilities of interaction between scientists and their data. The application offers improvements in the ways the user inputs, organizes, visualizes, and stores data. This application also serves as a presentation platform, allowing the user to show findings within the environment in which he/she is working. By using the *Sequence Generator* scientists are deeply experimenting with visual thinking processes that will facilitate the production of hypotheses and results while studying patterns within the data.



This first screen shows: total area of the interface, 1. Input screen. 2. Empty visualizer screen. 3. One individual being entered. 4. Menus and filters. 5. Sliders for the individuals and the data entered. 6. Trajectory in which the data has been entered. 7. Save box. 8. Markers places to highlight certain parts of the visual area.

From: *Mapping science*.

The *Sequence Generator* opens possibilities for its user that were not available in the methods used originally for the analysis of the data related to the scientific research. To input the data becomes an easy task that combines graphical elements with other visual aspects such as color and form. Relationships between the elements and their behavior within the *Sequence Generator* facilitates the process of interaction with the data.

From: *Mapping science*.

the sequence generator / how it works

Opening the application with a clean screen allows the user to focus on inputting the data. The first screen only reveals the points resembling the parts of the bat's body and the two screens in which the different interactions take place. Because behavior is repetitive and the parts of the body that are touched are finite, the data presented and used in the *Sequence Generator* has been transformed into graphical elements, making possible to visualize and recognize patterns from the beginning.

The first important step in using the *Sequence Generator* is the entering of the behaviors so the sequences can be built. The user enters a code that matches numbers used during data gathering. Coincidence between the numbers is important since these numbers correspond to the numbers of the tags that each captured bat has. These tags are often found in subsequent investigations allowing the study of the lifespan, reproductive condition, and changes in body condition of the specific specimen.

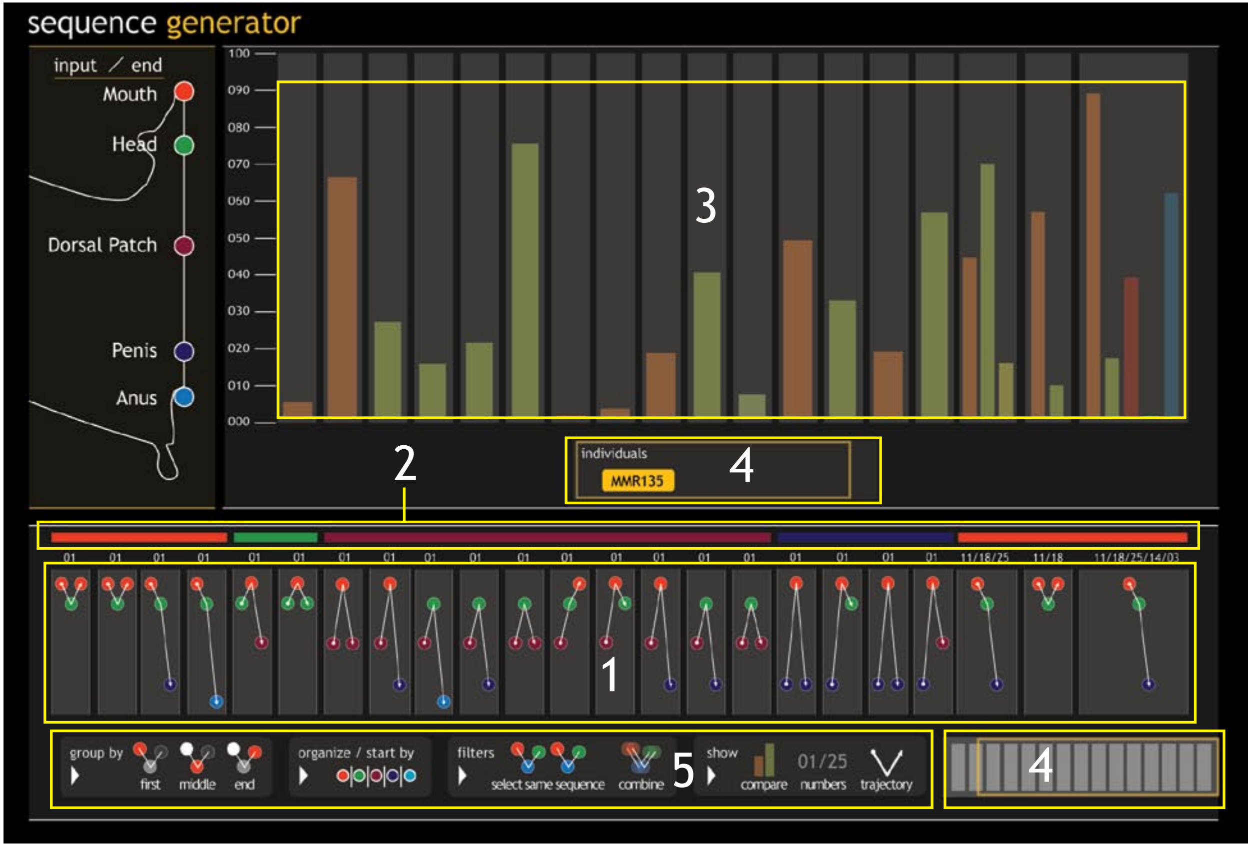
After entering the code, a grid that allows parts of the body to be input is displayed. This grid is not only for visual purposes but for algorithmic purposes as well, as applying a unique position to each part of the body greatly facilitates the construction of the sequences.

While entering data, the user can make corrections by selecting, deleting, or adding points. When parts of the body need to be entered to a new spot the add point tool adds a line in the grid to create a new spot in which the point can be entered.

Once the user has completed entering the data of an individual specimen, menus are displayed allowing analysis and interaction with the data itself. There are optional features included in the help tool menu, such as displaying the trajectory in which the parts of the body were entered, which facilitates de detection of patterns, or highlighting only certain fractions of the data. Other properties of the application are aimed to the manipulation of the data entered.

The user has an option of applying in and out markers, so parts of the data become highlighted visually. Positioning many markers allow to highlight more than one area and overlay the selection tool to determine connections between those selections. In the event that certain portions of the data go beyond the limits of the screen, sliders become visible.

Once the individual specimen's data has been analyzed, the user can click and drag any point to the sequence screen, where the sequences are displayed. All the sequences are built individually from the data entered in the input screen, representing every possible combination of parts of the body touched by the male bat. Numbers appear on top of each sequence to determine how many times that same sequence was repeated. This is one of the instances in the *Sequence Generator* in which numerical data complements the visual information presented.



This screen shows:
1. Visualizer screen showing several individuals.
2. Color coded bars to aid the user to recognize the groups. 3. Comparisons between individuals displayed. 4. Sliders for individuals and sequences. 5. Menus or options for select, group, organize, combine, and display data.

From: *Mapping science*.

The Sequence Generator features organization methods that transcend the original interaction with the data and offers additional display options. According to the scientist studying bats, these new options gave her the opportunity of exploring the data from new perspectives. In earlier explorations of the data, aspects like arranging the sequences was a fixated process with only one option. The organizational tools present in the Sequence Generator allow the scientist to display sequences following different criteria.

From: *Mapping science*.

A set of menus was developed responding to the scientist's needs. Since the scientist does not need individual repetitions while comparing the sequences, filtering and grouping options ultimately define the diagram. Therefore, the diagram adapts to fit to the user's primary interest, which is to study repetitions and frequency of the behavior performed by the bats.

Data sequences can be grouped so they are displayed following simple rules, such as all the sequences that start with the same part of the body, the ones that match the middle part, or the end part. As understanding the behaviors bears a strong relation with the interaction that the user experiences with the data, creating a structure for the displayed data allows interaction to be friendlier when the user is seeking to switch between parts of the body. The more options a user has at the moment of sorting the graphical elements the stronger his/her understanding of the data will be.

Sequences are usually focused on its middle section but not limited to it. Being able to focus on different parts of the sequences gives a broader idea of what the relationships are between the three parts that conform them.

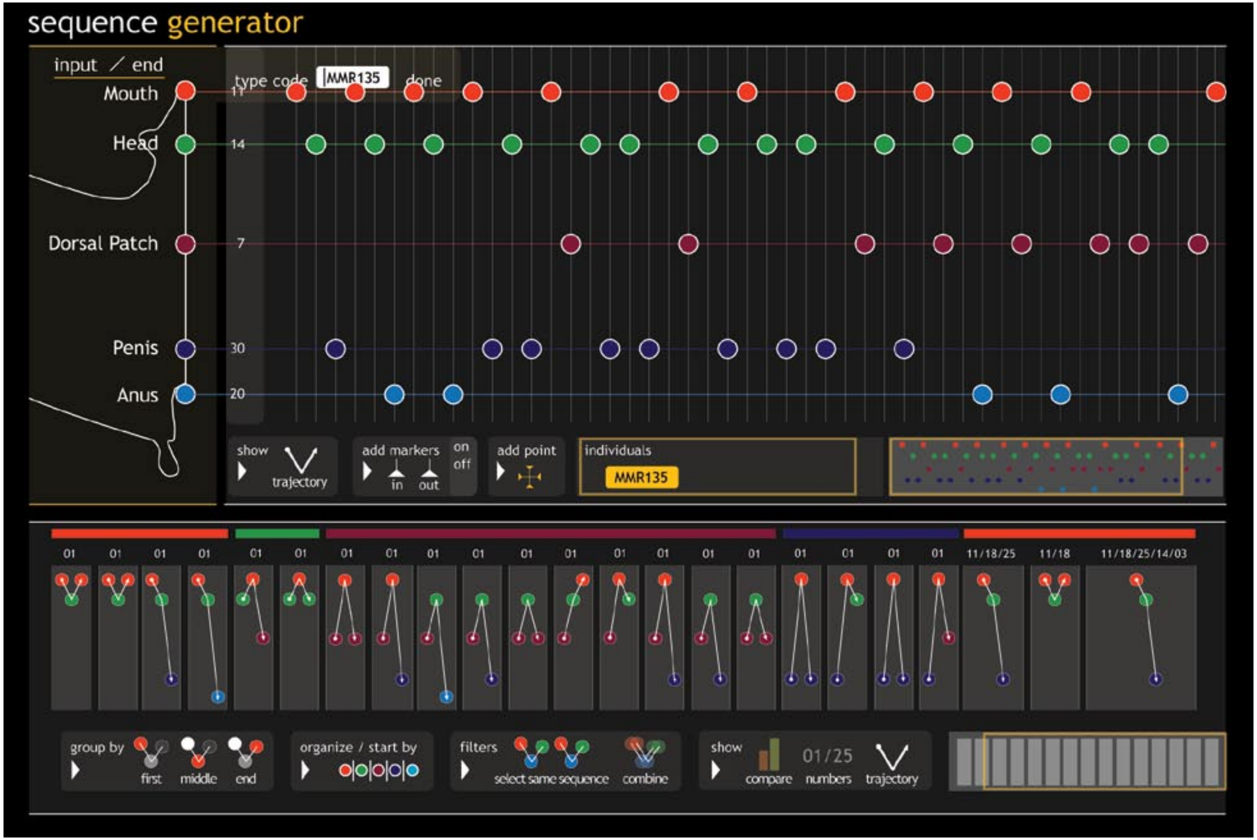
In order to improve the perception of the groups, colored bars are located above each group to indicate what part of the body the groups belong to. By default, the sequences display following the natural vertical order of the parts of the body in the bat.

If the user needs to reposition the groups so the diagram begins with other parts of the body (e.g., the head instead of the mouse), or connections between different parts of the body needed to be visualized (e.g., head and anus), the organize menu allows to do so.

Combining sequences is an important step in producing the final diagram, and by clicking on the select same sequence menu, the *Sequence Generator* will automatically highlight the sequences that are repeated. There will be a number displayed at the top of each sequence indicating how many times that sequence is repeated.

Several individuals can be combined or displayed in the same diagram; individuals can be inputted and then dragged into the sequence screen. When the user combines more than one individual the width of the sequence changes to adapt to the secondary information to be displayed when comparing individuals.

To define the connection within the graphical and numerical data is important but unnecessary when studying patterns between sequences. To facilitate this, the show menu switches information inherent to the sequences on and off. This menu displays numbers related to the repetitions of the sequences, graphs giving percentages for those repetitions, and the trajectory that shows connections between the three parts of a sequence.



Sequences and individuals can be visualized at the same time. Menus are displayed when something is dragged to the screens or input.

Sequences and individuals are available for further revisions. The *Sequence Generator* was designed with simplicity in mind so it does not represent a challenge for its user. To make it easy to learn and to understand are key to allow scientists to start using visual processes to interact with their data.

From: *Mapping science*.

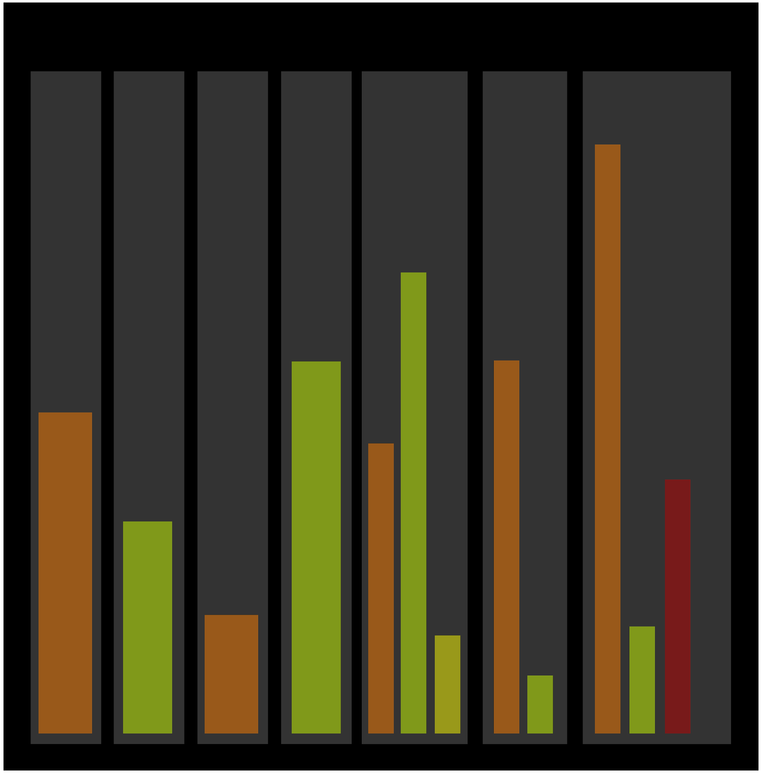
Menus allow the user to filter the data at any time. Different stages of the interaction get affected by applying the filters to the data. Input data can be modified, sequences can be combined, different parts of the body are arranged by groups, individuals get revised, raw data is compared with sequences, and trajectory is displayed.

From: *Mapping science*.

The comparison tool is closely related to the final diagram, once all the sequences have been combined. This submenu displays bar graphs comparing in percentages the repetitions of the behaviors performed by the male individuals. These bars help to quickly examine and establish a comparison between different individuals.

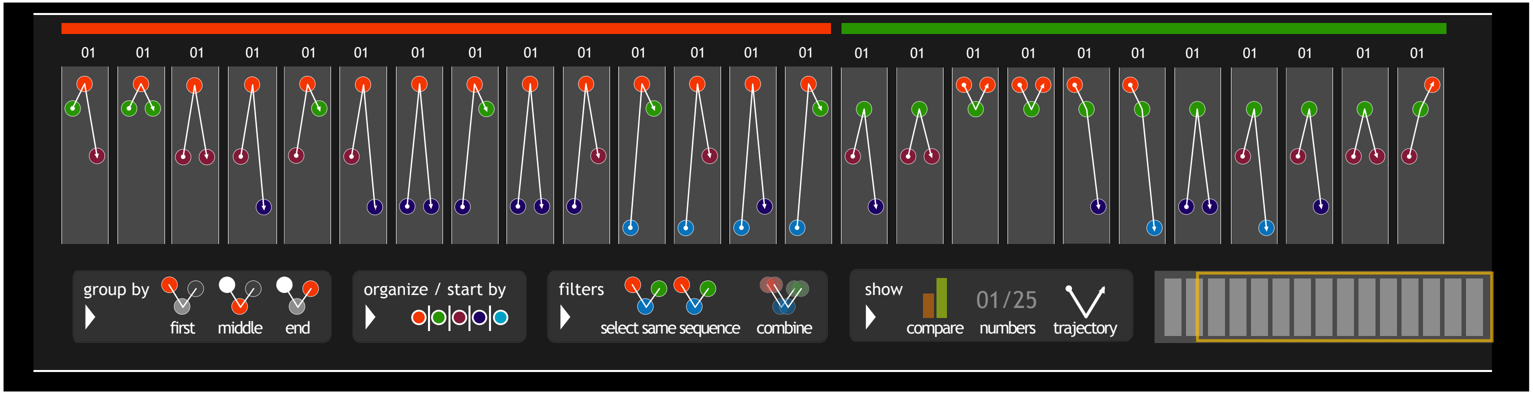
Individuals entered in the *Sequence Generator* are kept in a secondary database allowing their storing and retrieval for future references. However, the storing function of this tool needs to further developed.

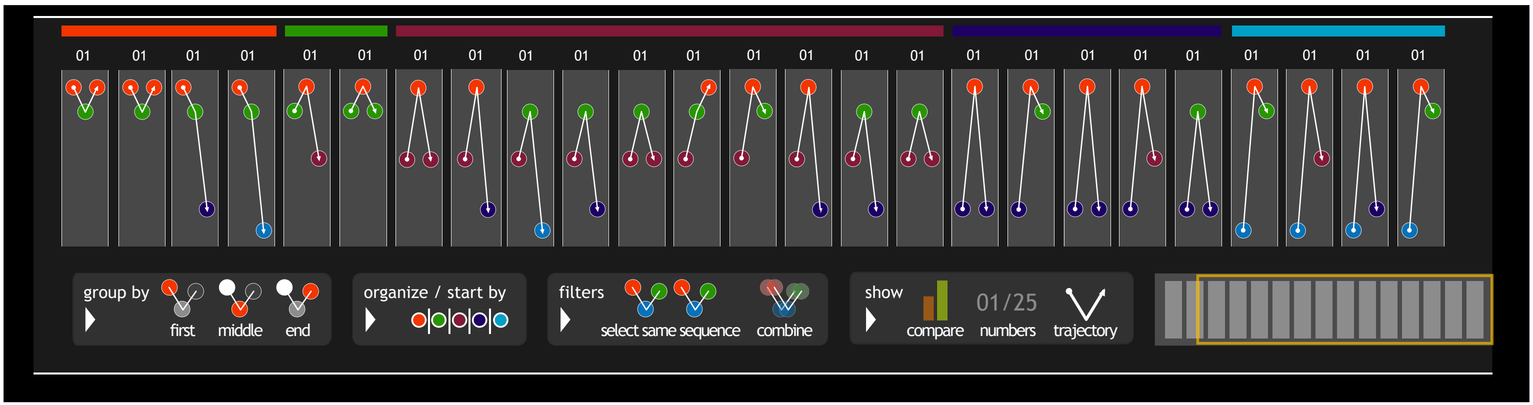
After the user has acquired proficiency in the use of the *Sequence Generator* some visual aspects will become iconic and the processing of the data turns into a more intuitive activity.



Comparisons between individuals allow the user to define what sequences are the most common ones, so patterns become more evident. In here, bars display one individual or several compared to each other. The width of the bars changes to facilitate recognition of more than one individual.

From: *Mapping science*.





In this two images, options for displaying the data can be seen. One of the strongest applications of the *Sequence Generator* is the *"group by filter"* which arranges sequences by their parts. Displaying the sequences by first, middle, or last element enhance the possibilities of finding patterns within the data.

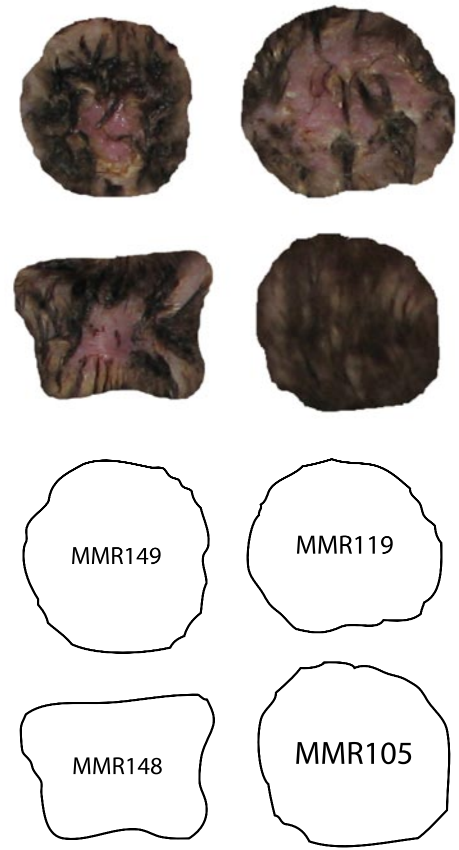
From: *Mapping science*.

case study # 2 / shapes

Equally important is the study of morphological characteristics of the dorsal patch. Due to its importance in defining mating choice and sexual selection, the patch's form and symmetry are key for the scientist to understand the reproductive function of bats.

The visual potentials of this part of the project were particularly appealing, as the information was not only translated, but rather improved instead by creating visualizations that allow to present the existing material in a clearer and cleaner way. Shapes were redrawn and digitalized to vectors to make them more malleable and in accordance to the medium.

The challenges of this case study were the development of organizational methods through which represent the data, finding structures for the data that needed to be displayed and built upon, and creating the environment for the interaction between the scientist and the database. Even though the scientist had to deal with dorsal patch shapes before the development of this tool, the digitization of the shapes allowed to experience a totally new approach. Scientists are familiar with the use of form, specially biologists, but the process that leads to the manipulation of those forms and the interpretation of the information that they carry are two different things.



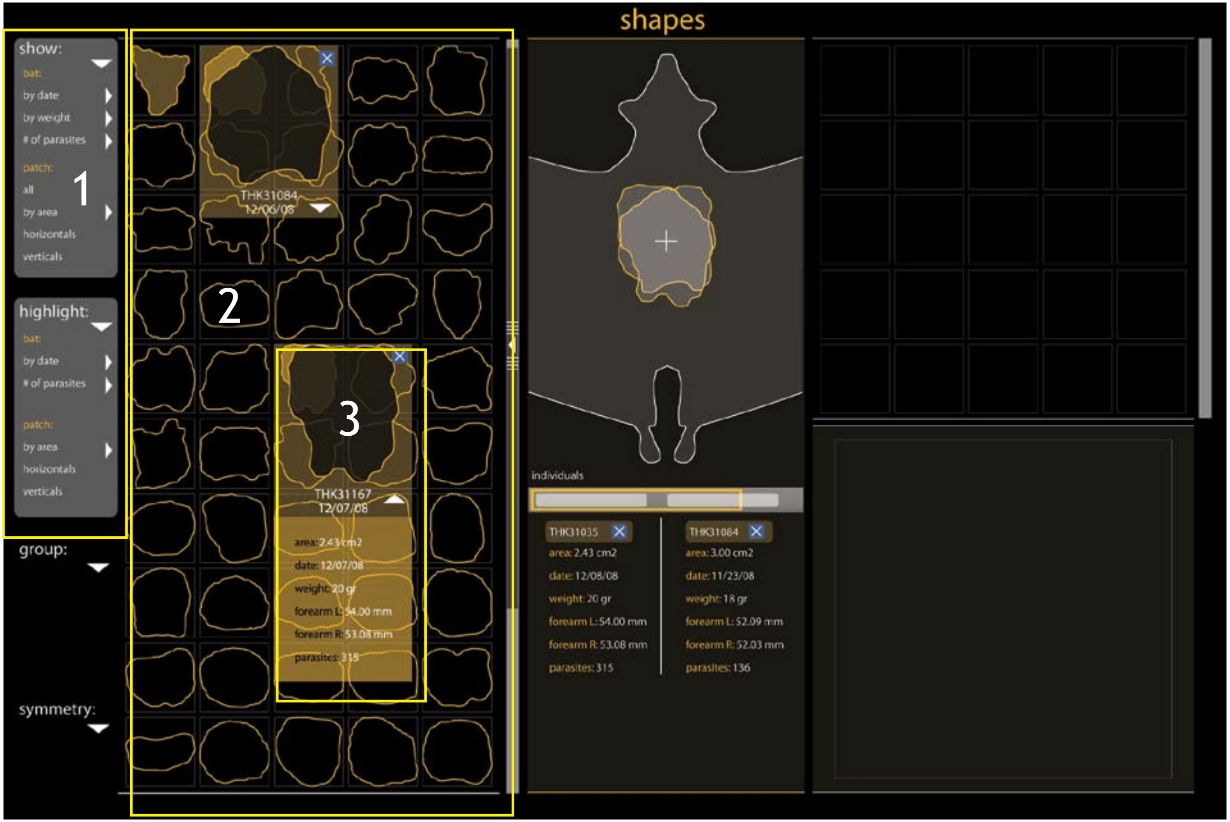
Original methods for the manipulation, storing, editing, and organization of the existing data were once again important triggers in defining the need for improvement. The scientist originally stored pictures of the bats creating new layers for the patches and having to open them individually when in need for comparison. An additional inconvenience was that the images were also heavy and lacked the resolution the analysis required. The application of design principles aimed to facilitate the medium for improvements in the performance of the scientific research.

Shapes utilizes a different process than the *Sequence Generator*, as visualizations are applied in an external level and then inputted to the database so the tool can retrieve them. Conversely, in the *Sequence Generator* the sequences are built within the tool, produced by the tool and organized by the tool. Although *Shapes* also deal with organization of date, the method utilized to interact with the data follows different rules.

The principles for the interaction in *Shapes* are based in the visual comparison of graphical representations and the ability to regroup them allowing the users to create their own categorization of the patterns found by analyzing the data. Perhaps one of the strongest features present in *Shapes* is the possibility to access both types of data, the numeric and the graphical representations. This feature relies on the connection that links the study of the forms with other data related to the patch, such as number of parasites, date gathered, or weight of the individual.

Both, the purely quantitative and the qualitative data. Shapes are combined with other data relative to the research so connections can be established.

From: *Mapping science*.

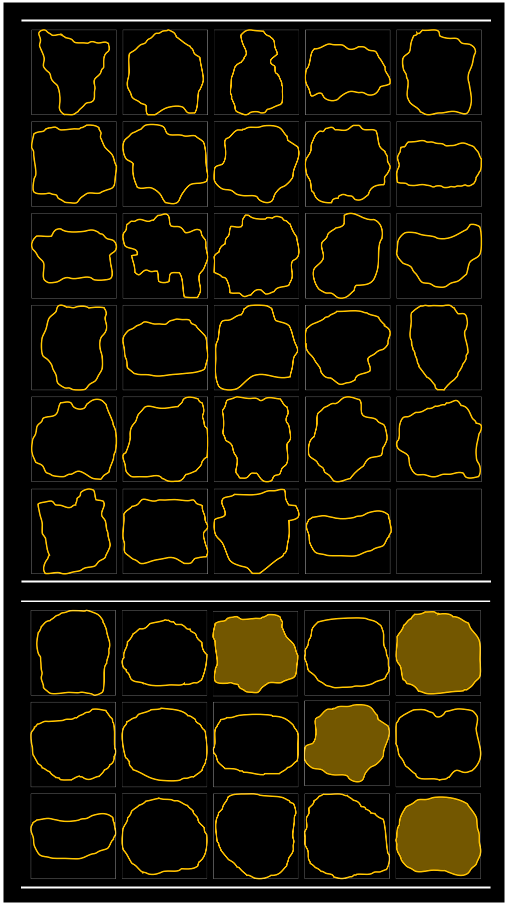


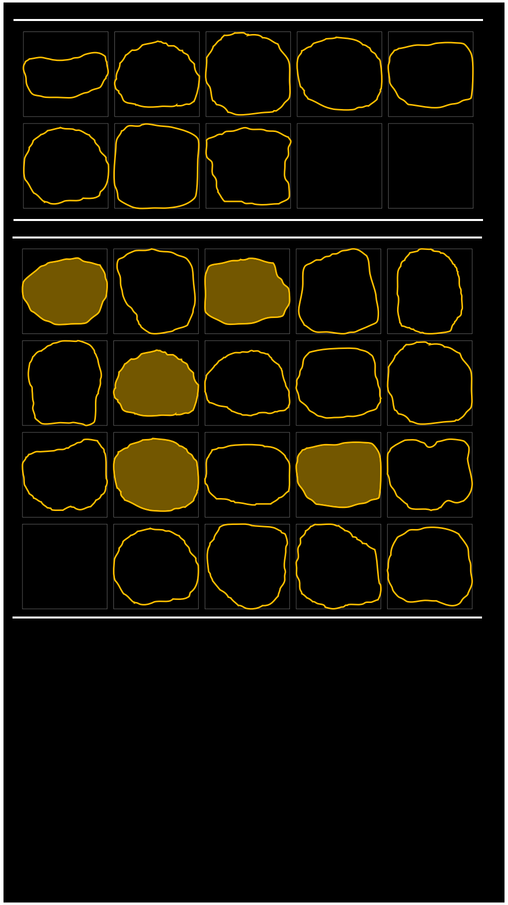
The left side of the screen visualizes the database and offers highlighting and organization possibilities of it. In the center, comparisons as well as original sizes of the shapes allow the user to connect visual cues with numerical data.
1. Menus and filters that show or highlight the data to access diferent characteristics of it. 2. The viewer screen displays selections of the data or the entire database. 3. Individual drop menus display additional quantitative data that can be compared with morphological traits.

From: *Mapping science*.

Organizing data is as important as displaying data. Being able to select what is displayed in the screen is a special feature in *Shapes*. The lack of categorization and hierarchy of the first attempts to organize the data created the need to develop this options in Shapes.

From: *Mapping science*.





Selecting and highlighting shapes can improve drastically the exploration of data. Commonalities between the shapes can become apparent to the eye when using visual representations of the shapes. These are just a few of the possible results of using the menus available in *Shapes*.

From: *Mapping science*.

shapes / how it works

When first using *Shapes* the user decides what information to display and how to display it. Using menus to retrieve information from the database allows the user to choose to display attributes related to the bats or the patches. Date, weight, and number of parasites are data that correspond to the individual specimens, while area, horizontal position and vertical position, refer directly to patch's attributes.

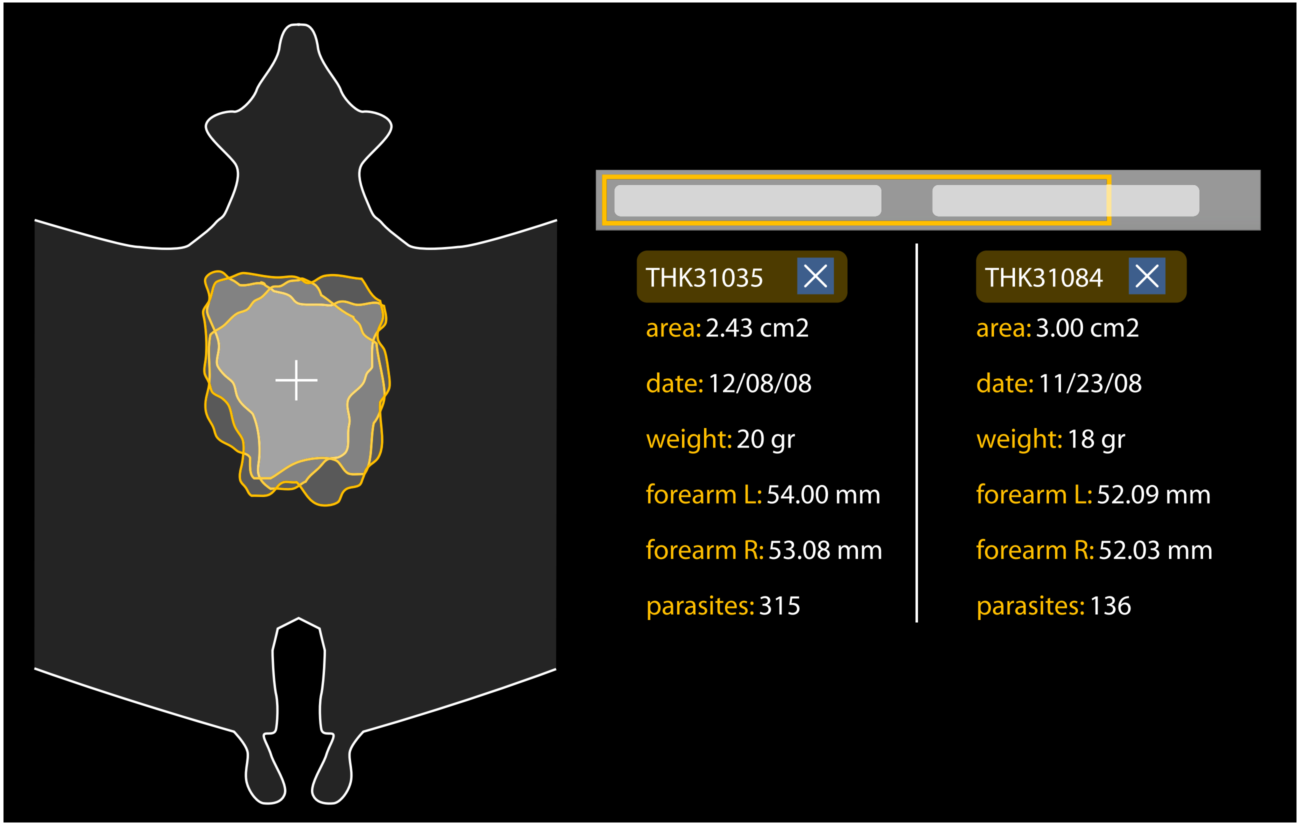
Following the display of the information, the user combines a series of filters in order to highlight certain characteristics of the data. Properties common to the shape and the numerical data can be displayed allowing the user to sort through well-known information. All the information is product of the scientist's research and this tool is aimed to fulfill the needs of the scientist for better ways to interact with his/her data. The scientist is in total awareness of the kind of data that he/she is manipulating, but not in control of the visual thinking processes that derived from this experience.

Features like the menu that allows finding dates in which data was gathered and the menu for number of parasites are solutions to explicit concerns of the scientist. For example, the importance of finding dates deals with the relationship that exists between the dates and the maturity or progress in the development of the patches. Some of the chemical compounds present in the patch are supposed to act as insect repellent and the symmetry of the patch is thought of as representative of the of the individual's health. Information used in *Shapes* keeps a close connection between all its parts; each element of data can be compared against other data, establishing connections that will bring insight to the process of analysis.

There are varied ways to access complementary information and the user can evaluate certain visual characteristics of the patches and visually display representations of the numerical data. The use of the show and highlight menus lays down visual cues. By placing the cursor over individual shapes the user can bring more detailed information of the patch and a complete list of the information available for that specific individual. A delay in the mouse-over avoids distractions while interacting with other segments of the interface.

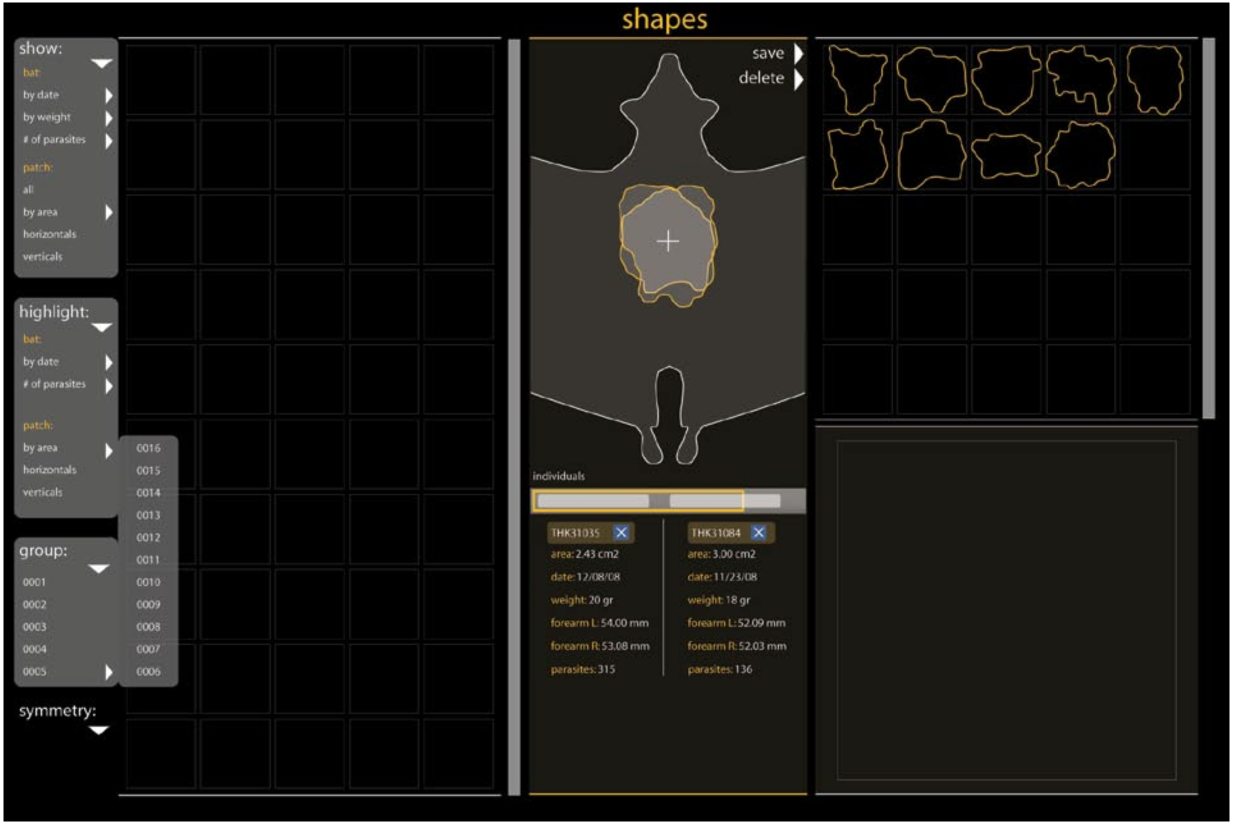
Original size of the shapes, spatial comparison, and the position of the patch within the bat, can be obtained by dragging a shape to a graphical synthesis of a bat located at the center of the interface. In this visualization, the patch in its original size is centered with respect to a reference point and available information displayed in the bottom of that section. A slider-like box shows the individuals displayed and allows the user to slide it to access information in case many shapes are being analyzed simultaneously.

Placing shapes together facilitates comparison against each other and morphological attributes such as area and smoothness of the shape can be easily contrasted using these visual elements. The option of hiding shapes or make them visible avoids visual noise when studying individual shapes but increases effectiveness while comparing several individuals.



Being able to compare shapes to each other and to other numerical data makes this tool even more manageable. *Shapes* avoid users to have to store one type of data in one place and other kinds in another. *Shapes* connects both types of data so they do not represent a visual or mental noise. The scientist is supposed feel comfortable comparing the qualitative and the quantitative data.

From: *Mapping science*.



The option of regrouping shapes facilitates the finding of new relationships between the data and improves the process of data mining.

From: *Mapping science*.

One of the primary goals of *Shapes* is to aid the user to find common aspects of the patches. Regrouping becomes paramount to this matter and gives users a way of creating subgroups that later will represent findings. When there are many groups saved, they can become another element to be compared.

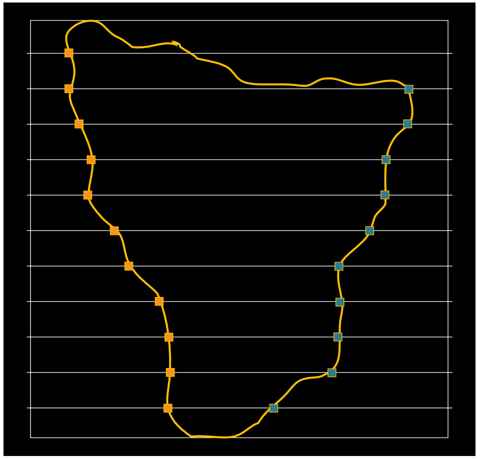
From: *Mapping science*.

Once the user has revised the shapes and contrasted their characteristics, there are two complementary tools that can be used. The top right box is the group tool, which allow the user to regroup shapes following rules of similarity. For example, when the user is looking for morphological aspects of the shapes, grouping the shapes focusing on their similarities with geometrical criteria could be helpful. Furthermore, when looking for relationships between dates and sizes of the patches, groping the shapes by the same date might contribute to the understanding of the relationships. To be more specific, older patches were found later in time and vice versa.

This aspect of *Shapes* combines visual thinking processes with analytical thinking, leading to a more effective production of hypothesis and theories regarding the scientific investigation. New groups often represent patterns within the shapes or within the data attached to them. Shapes from the left area of the interface can also be dragged to the grouping tool. Saved groups can be opened by accessing a menu or deleted when no longer needed.

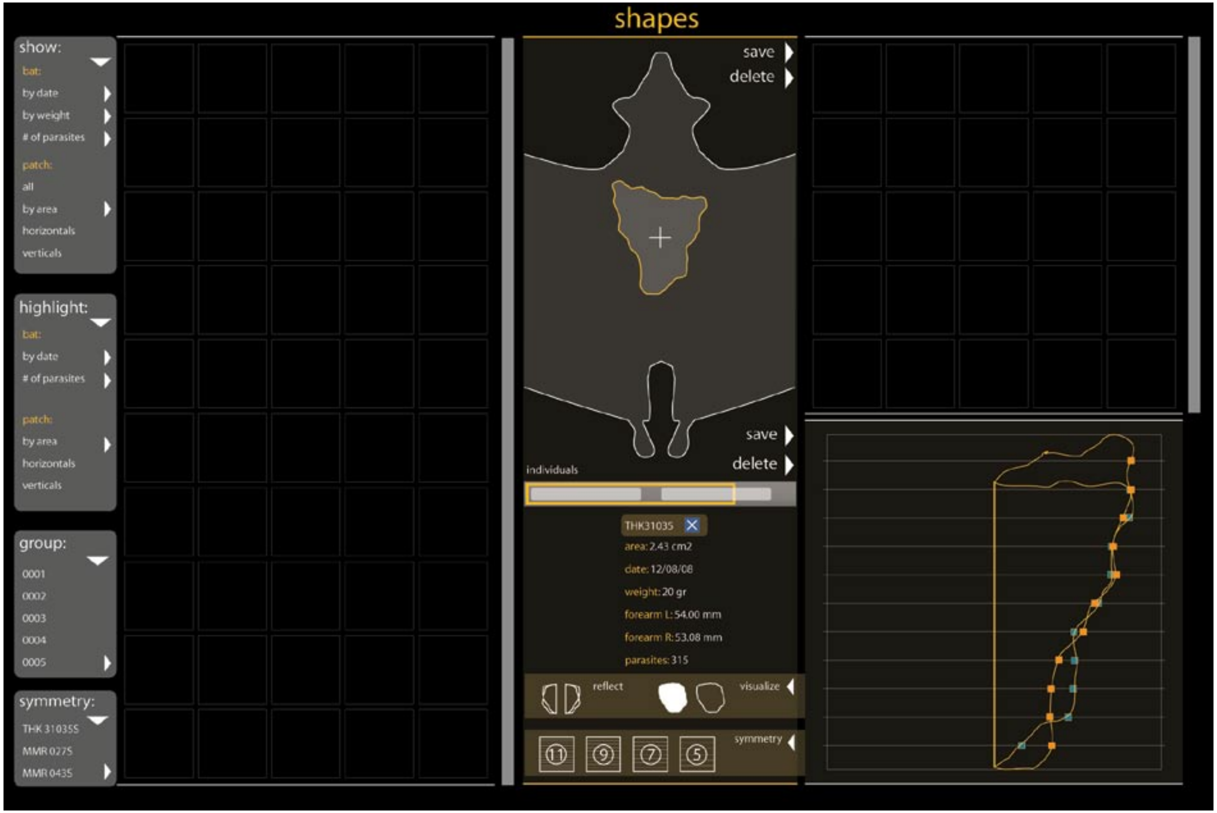
The second complementary tool (the symmetry tool) is perhaps one of the most important. This tool displays horizontal lines that allow the user to create landmarks by clicking the intersections of horizontal lines and the border of the shape. Symmetry is then measured by reflecting one side of the shape on top of the other and studying overlapping of landmarks. In some cases landmarks overlap perfectly to one another, which denotes a high level of symmetry. Reflecting the shape producing no overlapping between landmarks will demonstrate absence of symmetry.

The symmetry tool is primary to the morphological aspect of the investigation because its implication in defining if symmetry and sexual selection are correlated. Within *Shapes*, the symmetry tool is the only one that allows the user to modify the data. As well as groups, the result of applying the symmetry tool can be saved or deleted. The possibility to save groups and symmetries improves the database while interacting with the data, producing graphical representations of the findings and storing them create a more comprehensive collection of information related to the data.



The symmetry tool.

From: *Mapping science*.



The symmetry tool is one of the most important features of the interface, allowing the user to calculate and measure bilateral symmetry which is a key factor in the development of the research.

From: *Mapping science*.

Another tool that resulted from the interaction between scientist and designer while developing visual representations of the scientific research was the symmetry tool. Visual solutions created for general aspects of the patches translated to other applications like in the case of symmetry. Landmarks and reflection of the shapes would have been harder to ideate with the original pictures used to analyze the shapes.

From: *Mapping science*.

10 / CONCLUSION

conclusion

My thesis established a series of interactions between a scientist and a designer as the common ground for collaborative work in the production of scientific data visualizations for a Ph. D thesis on bats. The relationship between participants marks the success of the results in the production of such visualizations. The designer most understand not only the meaning of the content within the scientific data, but also the complex mechanism of implementing scientific problem-solving methods. Moreover, scientists need to evaluate and comprehend the role of visual thinking processes and visual thinkers in the visualization of their data.

By applying design principles and knowledge gathered from science and design, Mapping Science produced a set of prototypes that improved the conditions in which the scientist interacted with her data.

Implementation of visualization in the methodology of scientific research is key to the elaboration of more efficient systems. Mapping Science transcends the mere creation of specific visual solutions and investigated the use of these visual solutions within the process of exploring data in search for useful information carried by meaningless numbers.

By giving abstract characteristics to data through the means of visualization, the scientist in search for patterns within the data was able to pinpoint relationships and to establish connections between individual elements of that data, which facilitates creative resolution when producing hypotheses and theories related to the research. Visual images also provide better conditions for perception and mental processing of the content within the data.

Prototypes from the Mapping Science lab separated two core aspects of the Ph. D. thesis on bats, behavior and morphology. These two aspects were addressed by the creation of the interactive tools.

The first tool, the Sequence Generator, structures visual graphics to provide the scientist with advantages when entering, analyzing, and storing data. Visualizing behaviors from the beginning in a more graphical way enhances the possibilities of finding meaningful information.

The Sequence Generator displays behaviors entered by the scientist and integrates these behaviors in sequences, facilitating the reorganization and regrouping of the information. Methods such as sequential and non-sequential reorganization of the sequences are paramount to compare different parts of the data.

A second tool entitled Shapes deals with morphological aspects of the structure found in the bats, in this case the data is pre-processed first to create a database that provide the tool with the elements to be displayed. Shapes complements the search for patterns in form and symmetry creating a comprehensive collection of images that can be visualized, compared, and manipulated to find characterizations of the structure and different commonalities between the shapes and other quantitative information regarding the research. Numerical features are fundamental to establish connections with the theoretical part of the research.

Comparisons between the original methods used in the gathering and processing of the data and those supplied by the interactive tools developed in this thesis, show a remarkable improvement and state the advantages of using design-thinking processes in the systematic resolution of scientific researches. Both, the process of developing the visualizations and the interaction with the scientist in charge of the research allowed participants to construct a solid ground for future implementation of interactive visual tools in scientific problem-solving tasks.

Improvements to the interfaces are necessary since the production of these tools will demand testing and assessment of viability from other scientists familiar with the research or similar studies. Defining and testing design solutions is also needed, there are some aspects of interaction that can only be highlighted by the actual experience of using the tool.

Without disregarding the importance of improvements to be made, this thesis has fulfilled the primary concerns of scientist and designer, providing a strong knowledge and experience for future implementation of these tools into real research initiatives.

"If this book of mine fails to take a straight course, it is because I am lost in a strange region; I have no map."

Graham Greene, The End of The Affair

11 / REFERENCES

references

Davis, G. B., & Olson, M. H. (1985). Management Information Systems. McGraw Hill, New York.

Fry, B. (2008). Visualizing Data: Exploring and Explaining Data with the Processing Environment. O'Reilly Media, Inc., Sebastopol.

Hyrkäs, J. (2000). Scientific Visualization. Seminar of content creation, Helsinki University of Technology.

MacEachren, A. M. (2004). How Maps Work. Representation, Visualization, and Design. The Guilford Press, New York.

Polland, M. (1994). Scientific Visualization: A Synthesis of Historical Data. Department of Education, Technical report. 17p.

Richarz, K. and Limbrunner, A. (1993). The World of Bats. THF Publications, Neptune City.

Robin, H. (1992). The Scientific Image. From Cave to Computer. W. H. Freeman and Company, New York.

Schafersman, S. D. (1997). An Introduction to Science: Scientific Thinking and the Scientific Method. Electronic source: http://www.muohio.edu/˜schafesd/documents/intro-to-sci.html

Shedroff, N. (2000). Information Interaction Design: A Unified Field Theory of Design. Pp. 267 - 292, in Information Design (Jacobson, R., ed.). The MIT Press, Cambridge.

Southworth, M. and Southworth, S. (1982). Maps. A Visual Survey and Design Guide. Little, Brown and Company, Verona.

Tufte, E. R. (2005). Envisioning Information. Graphics Press, Cheshire.

Tufte, E. R. (2005). Visual Explanations. Images and Quantities, Evidence and Narrative. Graphics Press, Cheshire.

Tufte, E. R. (2006). The Visual Display of Quantitative Information. Graphics Press, Cheshire.

Ware, C. (2004). Information Visualization: Perception for Design. Morgan Kaufmann Series, Elsevier, San Francisco.

Woolman, M. (2002). Digital Information Graphics. Watson-Guptill Publications, New York.

Wurman, R. S. (2001). Information Anxiety 2. QUE, Indianapolis.

Zins, C. (2007). Conceptual Approaches for Defining Data, Information, and knoledge. Journal of the American Society for Information Science and Technology, 58(4): 479-493.